PROBLEMI DI ASTRONAUTICA E DI MECCANICA CELESTE

a ricordo del Professor Giuseppe Colombo

Atti del Simposio internazionale tenutosi presso il Politecnico di Torino
10-12 Giugno 1987

SUPPLEMENTO AL VOL. 122 (1988)
Atti della Accademia delle Scienze di Torino
Classe di Scienze Fisiche, Matematiche e Naturali

TORINO
ACCADEMIA DELLE SCIENZE
1990

PROBLEMI DI ASTRONAUTICA E DI MECCANICA CELESTE

a ricordo del Professor Giuseppe Colombo

Atti del Simposio internazionale tenutosi presso il Politecnico di Torino
10-12 giugno 1987

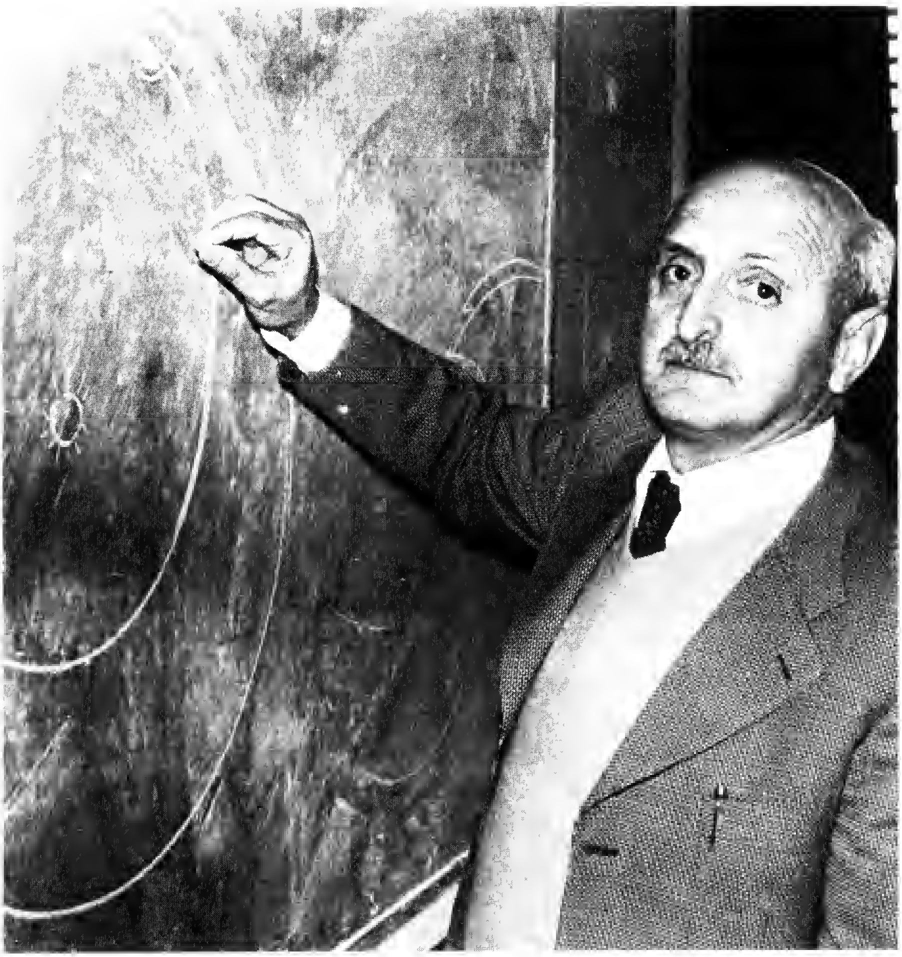
SUPPLEMENTO AL VOL. 122 (1988)
Atti della Accademia delle Scienze di Torino
Classe di Scienze Fisiche, Matematiche e Naturali

TORINO
ACCADEMIA DELLE SCIENZE
1990

Progetto scientifico e organizzazione della
ACCADEMIA DELLE SCIENZE DI TORINO

con il patrocino e la collaborazione di:

REGIONE PIEMONTE
CITTÀ DI TORINO
PROVINCIA DI TORINO
BANCO DI NAPOLI
CASSA DI RISPARMIO DI TORINO
FIAT AVIAZIONE



Giuseppe Colombo

3 Ottobre 1920 - 20 Febbraio 1984

SVOLGIMENTO DEI LAVORI E INDICE

12 giugno 1987

Presentazione

Questo volume raccoglie i testi delle relazioni presentate in occasione del Convegno tenutosi nel Giugno 1987 in onore di Giuseppe Colombo, immaturamente scomparso qualche anno prima. Non intendo qui ricordare le sue brillanti intuizioni scientifiche e i suggerimenti dai quali la ricerca spaziale trae ancora largo beneficio.

Questa Accademia, di cui Colombo era Socio, si prese il còmpito non solo di organizzare il Convegno ma anche di curare la stampa delle varie relazioni. Purtroppo, per eccesso di fiducia e difetto di esperienza, non fu posta come condizione categorica la consegna del manoscritto in forma definitiva all'atto della sua lettura.

La conseguenza inevitabile è stata che ci sono stati i ritardatari, i super-ritardatari, gli oberati da super-lavoro, i perfezionisti; ma anche, per fortuna, qualche buon cireneo, che qui ringraziamo. Da parte sua, l'Accademia, cronicamente a corto di personale, non ha nominato tempestivamente un Editor *che tirasse le fila dell'impresa.*

Molti autori tuttavia avevano consegnato puntualmente il testo delle loro relazioni, che vedono ora pubblicate con imperdonabile ritardo. È a loro che l'Accademia chiede scusa per mano dello scrivente, che questa funzione di Editor *se l'è presa a suo rischio e pericolo soltanto nel Maggio di quest'anno.*

L'augurio è che il ritardo non faccia velo alla qualità del contenuto.

Mario G. Fracastoro
Presidente
dell'Accademia delle Scienze

Torino, Settembre 1990.

Parole del Presidente dell'Accademia
Silvio Romano

Nell'aprire i lavori di questo simposio, rivolgo un deferente saluto alla memoria di Giuseppe Colombo, scomparso il 20 febbraio 1984 a Padova all'età di 63 anni: la sua è stata una morte prematura, che lo ha colto nel pieno vigore intellettuale, al culmine di una attività scientifica che per molti aspetti ha del prodigioso. Vincitore nel 1955 di un concorso per la cattedra di Meccanica razionale, dopo sei anni di insegnamento di detta materia presso le Università di Catania, di Modena e di Genova, veniva chiamato nel 1961 dalla Facoltà di Ingegneria di Padova come titolare di Meccanica delle vibrazioni: questa cattedra, per lui appositamente istituita, è stata la prima del genere in Italia.

Il 4 ottobre 1957 l'Unione Sovietica effettuava il lancio del primo satellite artificiale, il famoso Sputnik: sin dal giorno a cui risale tale avvenimento che segnò l'inizio dell'era spaziale, Colombo non rimase insensibile ed estraneo a tutta la mole di problemi che improvvisamente si presentavano e che rilanciavano come scienza di primissimo piano e di bruciante attualità la meccanica celeste. Colombo volge la sua attenzione a questa disciplina, sia nei suoi aspetti tradizionali riguardanti lo studio dei corpi celesti naturali, sia nei nuovi e più che mai complessi aspetti riguardanti lo studio del moto e della traiettoria dei missili e dei satelliti lanciati nello spazio e posti in orbita dall'uomo. Fin dal 1961, cioè fin dal giorno della sua chiamata alla cattedra di Meccanica delle vibrazioni presso la Facoltà di ingegneria di Padova, la Facoltà di scienze della stessa Università conferì a Colombo, che ormai si era affermato come uno dei massimi studiosi della materia, l'incarico dell'insegnamento della Meccanica celeste. A tale compito egli si dedicò con passione ed entusiasmo; e quando, a metà degli anni settanta, in applicazione di una assurda disposizione del nostro ordinamento universitario, l'incarico gli venne tolto, Colombo non rinunciò all'insegnamento della disciplina cui tanto aveva dato e andò ad impartire detto corso presso la Scuola Normale Superiore di Pisa, che lo aveva avuto come allievo tanti anni prima. Voglio anche ricordare che nel 1982, nella Facoltà di Ingegneria di Padova, Colombo era passato dalla cattedra di Meccanica delle vibrazioni a quella di Vettori e veicoli spaziali.

L'Accademia delle Scienze di Torino ha commemorato Giuseppe Colombo nella solenne riunione del 17 aprile 1985 con una commossa rievocazione della sua figura fatta, da par suo, da Dionigi Galletto. Giuseppe Colombo è uno di quegli uomini che rimangono vivi in quello che hanno dato alla scienza e la celebrazione di uno studioso di così alto livello non poteva concludersi con una commemorazione. Le celebrazioni, anche se guardano indietro verso il passato, servono a mobilitare le energie verso il futuro e la nostra Accademia ha ritenuto che il modo migliore per onorare un uomo che in molti campi ha accresciuto il patrimonio del sapere umano fosse quello di promuovere un incontro di insigni studiosi provenienti da diversi Paesi; un incontro che potesse divenire teatro di un dibattito vivo proprio in quei campi in cui il nostro illustre consocio scomparso ha svolto una attività feconda ed illuminata; un incontro che potesse concretarsi in un dialogo che, arricchito da una molteplicità di posizioni dottrinali, di esperienze e di metodi di indagine, contribuisse a quel progresso della scienza che è il nostro compito istituzionale.

Illustrato lo spirito con cui questo Simposio è stato ideato, io che opero in un campo di studi molto lontano dal vostro, non posso addentrarmi nel programma: *ne sutor ultra crepidam*. Credo di avere solo la veste per porgere a nome dell'Accademia delle Scienze di Torino, il più cordiale benvenuto alle Autorità la cui presenza conferisce alla riunione il miglior prestigio; lo stesso benvenuto do anche a tutti coloro che sono in questa sala. Ringrazio il Rettore del Politecnico di Torino per la sua cortese ospitalità: la nostra sede è al momento inagibile per gli importanti lavori di ristrutturazione che vi si stanno effettuando. Un deferente saluto rivolgo agli illustri rappresentanti delle Accademie consorelle qui affluiti: sono i rappresentanti della Accademia Nazionale dei Lincei, della Accademia Nazionale dei XL, dell'Accademia delle scienze dell'Istituto di Bologna, dell'Accademia Nazionale di scienze lettere ed arti di Modena, dell'Accademia Patavina di scienze lettere ed arti, dell'Istituto Lombardo Accademia di scienze e lettere, dell'istituto Veneto di scienze lettere ed arti. Il Presidente della Pontificia Accademia delle Scienze ha inviato il seguente telegramma: «La Pontificia Accademia delle Scienze è spiacente di non essere in grado di partecipare al Simposio internazionale su «Problemi di Astronautica e di Meccanica celeste». Ricordando grande figura professor Giuseppe Colombo eminente accademico pontificio qui commemorato ottobre 1986 occasione cinquantenario restaurazione nostra Accademia, invio congratulazioni e auguri partecipanti importante Convegno. Carlos Chagas».

Agli illustri Relatori che hanno accolto il nostro invito di offrire a

questo simposio le loro esperienze di pensiero, esprimo il vivo compiacimento e la riconoscenza più sincera; a loro è dovuta la ricchezza di questo incontro, ricchezza che non consiste soltanto nel misurare e nel comunicare progressi ritenuti acquisiti, ma anche nell'approfondire temi di studio in un ambito nel quale un continuo sviluppo ed un continuo arricchimento costituiscono quasi una necessità. Ringrazio il Comitato scientifico e la Segreteria amministrativa del Convegno, ringraziamento che estendo a quegli Enti di cui l'Accademia ha sentito attorno la solidarietà: sono la Regione Piemonte, la Città di Torino, la Provincia di Torino, il Banco di Napoli, la Cassa di Risparmio di Torino, la Fiat Aviazione.

Termino con l'augurio che i Vostri lavori siano il più possibile proficui: di ciò costituisce sicura garanzia il valore dei partecipanti.

Welcoming address

Franco BEVILACQUA *

Distinguished guests, ladies and gentlemen:

on behalf of the Center for Astrodynamics of Turin, of which I am the treasurer, let me welcome you to the present symposium.

Most of you are probably well acquainted with the scientific achievements of the late Professor Giuseppe Colombo, so that these achievements might not be reminded here. Yet a commemoration appears fair, inasmuch as his works have affected both the academic and the industrial community as well.

Giuseppe Colombo was born in Padua on October 2nd, 1920. After graduating in mathematics, he joined the teaching staff at the University of Padua. There he lectured on the mechanics of vibrations for twenty years since 1962. But he also took a deep interest in celestial mechanics, and in the early sixties made his first important contribution to this field.

At that time, the planet Mercury could be observed from Earth only with difficulty. Mercury was then believed to have the same period for both its revolution around the Sun and for rotation around its own axis. Colombo, on the contrary, while studying the resonances among celestial bodies, made the prediction that the rotation period of Mercury should be two thirds of its revolution period, namely 59 days for rotation versus 88 days for revolution. This prediction finally proved completely correct when the flybys of Mercury by means of spacecrafts enabled good pictures of the planet to be taken.

Another remarkable contribution by Colombo regarded the trajectory of the Mercury 10 probe. While NASA experts were planning an orbit leading to one or, at best, two Mercury flybys by the probe, Colombo suggested a modification of the orbit that would have enabled three of them to be made. Once again his prediction was correct, and, in 1974, the three flybys took indeed place.

Topics like the resonance among celestial bodies and the study of their stability charmed Colombo. Just after the discovery of the rings of Uranus, he was able to identify these rings on old plates taken before their

* Space Systems Group. Aeritalia, Turin.

discovery, and then made a discussion of the system's stability. His results, however, could be published only in part, and he always felt unhappy about that. This lack of time was certainly due to Colombo's deep involvment with the most important space missions set up by NASA for the exploration of the Solar System, such as the Pioneer 10 and 11, and the Voyager 1 and 2 missions.

The advent of the Shuttle transportation system offered Colombo the opportunity to put forward his best contributions to space science. In the seventies it was well known that a satellite could not be put into orbit through the most dense atmosphere layers, inasmuch as the friction would cause it to fall on Earth in a short time span. Colombo's idea to solve the problem was to hang the satellite up to something flying higher, that is the Space Shuttle itself. So the idea of the tethered satellited was born. The tether could have a suitable length ranging from some dozen to a few hundred kilometers. This circumstance, however, created a system for which the gravity gradient is no longer negligible. And so the idea was born to use tethers not just for the Shuttle itself, but for the coming space stations too. Tethered platforms, tethered constellations, and even tethers for use on the moon, are currently under consideration by the most important space agencies all over the world.

As for Colombo, in 1982 he was elected to the chair of «Space Vectors and Vehicles» that had just been established for him at the University of Padua. He also carried on his cooperation with NASA, and, at the Smithsonian Astrophysical Observatory, in Boston, he created a school that was soon to become a focal point in the study of tethered systems. In June 1983 Colombo was the only Italian scientist to ever be awarded the NASA medal for «exceptional results in space research». But now, suddenly, he passed over: cancer took him away on February 20, 1984, in Padua.

Even if Professor Colombo is no longer with us, let me say that his ideas are with us. The Tethered Satellite System is going to become a reality. In fact, an industrial agreement has been reached to build up a five hundred kilograms satellite to be brought into space by the Shuttle, and then deployed and retreived by means of a tether up to a hundred kilometers long. While the deployment and retrieval mechanism will be constructed by the Martin Marietta Corporation of Denver, Colorado, the satellite itself is presently under construction by Aeritalia Space Systems Group in this very town of Turin. And the town of Turin aims at becoming an international center for space studies by promoting all kinds of activity in the field. For instance, in 1985 a Center for Astro-

ter groups the representatives of both the academic and industrial world that are operating in this town. Next with the Academy of Sciences of Turin, the «Giuseppe Colombo» Center for Astrodynamics of Turin is most glad to welcome you at this international symposium, and wishes to all of you a very profitable experience over the next few days.

Thank you very much.

Il contributo di Giuseppe Colombo alla Meccanica Celeste

Giuseppe GRIOLI

Giuseppe Colombo, vinse, primo ternato nel 1955, il concorso per la cattedra di Meccanica razionale, disciplina che, come Egli stesso spesso dichiarava, costituì la base fondamentale di ogni Sua ricerca futura. Insegnò prima a Catania e successivamente a Modena e a Genova. Infine fece ritorno alla Sua Padova, chiamato nella facoltà di Ingegneria ove tenne l'insegnamento di Meccanica delle vibrazioni prima e successivamente quello di Vettori e veicoli spaziali. Insegnò anche Meccanica celeste a Padova e a Pisa. Dopo qualche anno, però, dovette lasciare l'insegnamento di tale disciplina a Padova a causa di quel capolavoro di oscurantismo della legge che subordinava l'assegnazione degli incarichi universitari a una certa graduatoria per fasce e il giudizio di studiosità a un'alzata di mano. Era in loco un esperto a livello mondiale ma gli studenti non potevano giovarsene perché la legge lo impediva! Essa avrebbe impedito di insegnare Relatività ad Einstein redivivo! Ne risentiamo ancora gli effetti.

Giuseppe Colombo tenne corsi e conferenze nelle principali istituzioni scientifiche di tutto il mondo ed ebbe notevoli riconoscimenti accademici. Fu membro dell'Accademia Nazionale dei Lincei, dell'Accademia pontificia delle Scienze, delle Accademie di Padova e di Venezia e anche di questa Accademia.

Conobbi Giuseppe Colombo nel lontano 1949 quando succedetti nell'insegnamento della Meccanica razionale a Ernesto Laura divenuto fuori ruolo. Il giovane Colombo divenne mio collaboratore in qualità di assistente. Il comune interesse verso i problemi della Meccanica razionale e della Fisica Matematica e le qualità umane generarono tra noi rapporti di amichevole stima e di amicizia.

Ovunque in Italia e fuori è stato degnamente ricordato il contributo che Egli ha dato alla Meccanica celeste e in particolare allo studio della dinamica del sistema solare. È naturale. Egli fu un vero Maestro in tale campo e i suoi studi, a differenza di altri pure di notevole interesse, hanno per oggetto il più affascinante dei problemi, quello di indagare i misteri dello spazio, di comprendere dove ci troviamo.

Attratto Egli potentemente da simili misteri, a un certo momento della vita i Suoi interessi scientifichi si orientarono completamente verso lo studio matematico e fisico matematico del Cosmo, soprattutto verso i problemi del sistema solare e in particolare dello stesso pianeta sul quale ci troviamo.

Taluni risultati, esposti anche in modo divulgativo ma rigoroso in conferenze varie, interessano anche il grande pubblico attratto dalla suggestione dell'argomento ma la vita scientifica di Giuseppe Colombo ha avuto inizio ancor prima nel campo della Matematica pura e della Fisica Matematica pura che, per Sua stessa dichiarazione, costituiscono l'indispensabile background del periodo successivo, quello dedicato alla Meccanica celeste. In tali studi l'acutezza di taluni procedimenti, la capacità di trovare la via adatta non sono affatto inferiori a quelle mostrate in seguito anzi, io credo, a volte superiori ma essi sono meno conosciuti perché interessano solo i cultori di Matematica, Fisica matematica e Fisica teorica e pertanto non hanno avuto risonanza al di fuori degli ambienti specialistici.

Le ricerche del primo periodo costituiscono la base di quelle del secondo e per rendersi conto della personalità scientifica di Giuseppe Colombo e comprendere bene l'intero arco della Sua vita di ricercatore conviene darne un breve cenno.

Un gruppo di ricerche riguarda la Meccanica dei continui (sistemi a infiniti gradi di libertà). Esse concernono le vibrazioni di una sfera immersa in un fluido, le piccole oscillazioni di una superficie conica, talune limitazioni superiori nel problema dell'equilibrio elastico atte a fornire condizioni sufficienti di sicurezza nella Scienza delle costruzioni, l'isteresi. Più numerose ricerche si riferiscono ai sistemi a un numero finito di gradi di libertà: dinamica del binario, moto di un corpuscolo elettrizzato in presenza di un dipolo, la dinamica dei corpi rigidi, la stabilità lineare e non lineare, un sottile studio sull'accoppiamento di due oscillatori, le oscillazioni non lineari, la teoria delle orbite.

Da rimarcare in questo primo periodo della vita scientifica di Giuseppe Colombo un sottile studio sull'eventuale influenza di termini di ordine superiore della soluzione principale sulla sua stabilità lineare, come pure una ricerca sull'esistenza di moti asintoticamente stabili nella dinamica di un sistema autonomo e una sulla determinazione di condizioni sufficienti per l'esistenza di orbite periodiche per un sistema conservativo a due gradi di libertà.

Queste ricerche indicano un evolversi degli interessi scientifici di Giuseppe Colombo verso questioni che sono di ricerca pura ma che già tendono a formare un vero rieercatore di Meccanica celeste.

Dopo un primo periodo di studi dedicato alla Matematica pura e alla Fisica matematica pura l'attività scientifica di Giuseppe Colombo ha una svolta singolare: solitamente un giovane si immette sin dall'inizio in un sentiero della Scienza che gli è congeniale e questo rimane poi la sua strada principale. Invece Egli che pure aveva colto risultati non peregrini, abbandona le ricerche di pura Matematica e volge ogni Suo interesse verso i problemi concreti della Meccanica celeste, con particolare attenzione alla dinamica del sistema solare. Egli stesso lo dichiara nel discorso pronunciato nel 1983 in occasione dell'adunanza solenne dell'Istituto Veneto di Scienze, Lettere ed Arti nella Sala dello Scrutinio in Palazzo Ducale in Venezia ove, testualmente, dice: «... Passando dalla matematica all'ingegneria, io sono diventato in questi ultimi 25 anni un tecnico specializzato. Contribuendo al progresso della ricerca spaziale, in particolare nel campo dell'esplorazione dei pianeti e della terra come uno di essi, ma anche nel campo applicativo, penso ormai di sapere abbastanza bene cosa è successo e anche di poter dare una ragionevole idea di quello che potrebbe essere il suo futuro ...».

Nell'immettersi nel suggestivo campo della Meccanica celeste, Giuseppe Colombo compie come primo passo quello più naturale, dato il particolare e profondo background acquisito nel precedente periodo di studi, e rivolge innanzitutto la Sua attenzione a questioni che si giovano della dinamica dei corpi rigidi, della teoria della stabilità, della teoria delle orbite, di quella delle equazioni differenziali. Sono argomenti di cui Egli si era occupato con successo in precedenza e che costituiscono una base fondamentale per i nuovi studi.

Come mai una tale svolta? Forse da tempo nel Suo animo covava una propensione verso lo studio dei problemi della Natura ma episodi significativi della Sua vita hanno certo influito sulla scelta definitiva. Mi piace ricordare, a tale proposito, quanto Egli dichiarò nel 1967 nella conferenza tenuta al convegno UMI. Egli disse, testualmente: «Quasi 20 anni fa, in questa sede, il prof. G. Krall teneva una conferenza dal titolo: «Mete lontane nel moto dei sistemi dinamici». Io ebbi il piacere di leggerla nel fascicolo della rivista Scienza e Tecnica dell'Università di Trieste e rimasi impressionato, oltre che dalla chiarezza di esposizione, dall'affascinante argomento». Io penso che certamente sul Suo orientamento avrà influito anche quell'evento straordinario che nel 1957 fu il lancio del primo sputnik per opera dei Russi, avvenimento che apriva all'umanità l'era spaziale.

Colombo ben sapeva che nelle prime decadi del nostro secolo si era avuto un decadimento degli studi di Meccanica celeste, in particolare di quelli riguardanti il nostro sistema planetario. Gli studiosi, specie i

giovani, si erano rivolti in massa verso lo studio dell'atomo e del suo nucleo, trascurando settori altrettanto fondamentali della Scienza. Forse oggi si ha una situazione analoga, pericolosa in prospettiva, in relazione all'informatica e alla teoria e all'uso dei computers. Essi costituiscono indubbiamente validissimi e insostituibili supporti della ricerca scientifica e della tecnica ma tuttavia distolgono molti, forse troppi, da settori altrettanto interessanti della Scienza, settori da cui dovranno provenire i problemi alla matematica computazionale e all'informatica.

La teoria delle orbite e il problema dei tre corpi fu poco studiata nella prima parte del nostro secolo, se si eccettua l'attività di un gruppo di matematici russi. Forse, ciò spiega come il primo successo spaziale fu colto proprio dai Russi con il lancio del primo sputnik. Subito dopo, quasi ovunque, in particolare in America, si ebbe un rifiorire di discipline prima trascurate quali ad es. la Meccanica celeste il cui studio oggi è fuori discussione,.

Innumerevoli furono le questioni di cui si occupò Colombo in una prima parte di ricerche più direttamente collegate alla Sua attività precedente di matematico puro. Egli considerò questioni inerenti i satelliti artificiali (1961-1965) studiando, in particolare, il problema ristretto dei 2, 3, 4 corpi, considerando, cioé, il satellite in presenza della sola Terra, della Terra e della Luna, della Terra, della Luna e del Sole e ritenendo la sua massa ininfluente sul moto degli altri corpi. In tal tipo di problemi Colombo si interessò di traiettorie, della stabilità di soluzioni significative, dell'influenza della temperatura che causando una deformazione del satellite ne fa variare i momenti di inerzia e conseguente influisce sulla sua dinamica, dell'influenza di coppie magnetiche.

Altri studi, più propriamente volti verso la dinamica del sistema solare, riguardano le leggi di Cassini, i moti di rotazione propria dei pianeti Mercurio (1965) e Venere (1967), l'influenza del Sole sullo spin di Mercurio, la cattura di polvere cosmica da parte della Terra (1965), la dinamica di piccole particelle nel sistema solare (1965), la formazione dei satelliti esterni di Giove (1971), l'influenza della cintura di asteroidi sulla storia del sistema solare, la dinamica dei Saturno e dei suoi anelli a proposito dei quali viene studiata l'ipotesi di una struttura a spirale onde spiegare l'asimmetria di lucentezza (1976), lo schiacciamento del pianeta Urano (1980).

Trattasi di una vastissima produzione scientifica, densa di richiami bibiografici e sempre dotata di un'accurata messa a punto del quadro della situazione attuale. Sarebbe lungo entrare nei dettagli; mi soffermerò brevemente su qualche ricerca maggiormente significativa.

Nello studio estremamente accurato del moto di un satellite del sistema Terra Luna, Colombo si pone il problema della determinazione di una possibile orbita per un corpo lanciato da una posizione vicina alla Terra che dopo alcuni giri intorno alla Luna torna sulla Terra ma supponendo ellittica la traiettoria della Luna, a differenza di quanto aveva fatto prima A. Jegorov che l'aveva supposta circolare. Si mostra, tra l'altro, che le note soluzioni collineari del problema dei tre corpi note quando si supponga trascurabile la massa di uno di essi sussistono anche quando si ritenga che i tre corpi abbiano tutti e tre massa finita. In tale studio non mancano considerazioni sulla stabilità della soluzione. In particolare, viene osservato come una debolissima azione stabilizzatrice può mantenere una stazione spaziale in una posizione opportunamente scelta tra quelle di una configurazione collineare.

Notevole lo studio dinamico del moto della Terra attorno al suo baricentro. Si discute innanzitutto su cosa si debba intendere per «sistema Terra»: esso arriva sino a una certa quota di atmosfera? Comprende gli oceani? Va tenuto conto della deformazione della crosta terrestre? Quale deve essere il riferimento per la valutazione del moto? È evidente che le risposte a tali quesiti influenzano in modo determinante il modello matematico e le sue difficoltà analitiche. Un'accurata discussione critica segue la precisazione delle equazioni dinamiche, soprattutto in relazione all'ordine di grandezza dei coefficienti significativi, con lo scopo di giungere a uno schema matematico che consenta una concreta valutazione numerica. I risultati raggiunti vengono ragionevolmente ritenuti come una prima approssimazione della realtà e vengono segnalati i difficili problemi teorici e di osservazione ancora aperti per una più precisa trattazione dell'interessante na difficile problema.

In questo studio è manifesto il prezioso ausilio del precedente background matematico, in particolare quello relativo alla teoria delle oscillazioni non lineari.

Di notevole interesse uno studio sull'evoluzione del sistema solare (1970). Colombo ripete con Alfvén: «In the beginning was the plasma» e osserva come non sia facile stabilire ove termina la fase formativa del sistema solare e ove inizia quella evolutiva. Forse è ragionevole ritenere che la fase iniziale della formazione sia terminata e abbia avuto inizio quella evolutiva attuale proprio quando sono rimaste predominanti solo le forze gravitazionali. In realtà una delle maggiori difficoltà per lo studio evolutivo del sistema solare è dovuta alla scarsa conoscenza della sua storia e delle sue origini.

Con particolare riferimento a Terra, Luna e Sole, Colombo svolge interessanti considerazioni sugli effetti di marea e su quelli dovuti a col-

lisioni di asteroidi ai quali Egli attribuisce la presenza di innumerevoli crateri visibili su Venere, Mercurio e Marte.

In questi studi è palese la commozione che accompagna Colombo quando si accosta ai problemi dell'Universo. Spesso Egli la manifesta apertamente come quando paragona le configurazioni celesti alla geometria dei cristalli e dei fiori e lega l'intero sistema solare ai meravigliosi fenomeni che ci è dato osservare sulla Terra. Egli osserva che ci si forma un'idea di come tutto evolve semplicemente stando seduti in piazza S. Marco a Venezia mentre si ascolta un quartetto e si assapora un caffé italiano (quando dichiarava ciò Egli trovavasi in America), osservando il luccichio delle stelle e le oscillazioni del mare che lambisce il molo a volte superandolo.

Notevoli ricerche sono quelle sul moto di rotazione propria del pianeta Mercurio che hanno procurato al loro Autore notevole notorietà. La questione della durata della rotazione propria del pianeta è stata oggetto della considerazione di molti studiosi: Schöder, Schiapparelli, Massa, Dollfus, Peale e Gold e altri ancora. Per parecchio tempo si è ritenuto vero un periodo di 88 giorni ma successivamente l'affinarsi delle osservazioni ottiche della superficie del pianeta ha mostrato che quel valore non era ammissibile. Osservazioni di Pettengrille e Dyce (1965) inducevano a ridurre quel valore e a portarlo vicino ai 60 giorni. È stato Colombo a proporre — in una lettera inviata a Nature nel 1965 — come consistente con l'osservazione ottica il valore di 2/3 del periodo orbitale. L'induzione di Colombo si fondava sull'osservazione che, tenuto conto dell'azione solare e di una certa differenza tra i valori massimo e minimo dei momenti d'inerzia equatoriali, al valore del periodo di rotazione propria uguale ai 2/3 di quello orbitale corrisponde una rotazione propria stabile. Insieme a Shapiro Egli ha creato un accurato modello matematico in cui si dimostra appunto l'esistenza di una tale soluzione stabile. I due Autori non mancano di indicare i molti e complessi problemi ancora aperti sull'argomento anche in relazione alla storia dell'intero sistema planetario e alla sua evoluzione.

Aceanto a una copiosa messe di studi teorici sulla dinamica del sistema solare va considerato un altro insieme di ricerche volto a mostrare la realizzabilità di missioni spaziali e la messa in orbita di laboratori per osservazioni del Cosmo, dei pianeti, dell'atmosfera, della stessa Terra e per misure di gravità, di campi elettromagnetici, ecc..

Ricorderò uno studio sulla possibilità di costruire un'orbita che permetta a una sonda di passare per ben tre volte nelle vicinanze di Mercurio anziché due come si era ritenuto possibile in base a studi di ricercatori

americani, con evidente crescita della resa di lancio del 50%. Ricorderò pure la dimostrazione di fattibilità di una missione su Giove con un volo di 400-500 giorni e della sua convenienza scientifica.

Contributi decisivi furono dati da Colombo alla realizzazione del progetto europeo per l'invio di una sonda in prossimità della cometa di Halley che, com'è ben noto, due anni fa passò alla minima distanza dalla Terra, denominato progetto Giotto dal nome del celebre pittore che, pare, abbia raffigurato la cometa in uno dei suoi dipinti nella cappella degli Scrovegni in Padova.

Alla fine del secolo scorso, nel 1895, il russo Kostantin Tsiolkovsky aveva avuto la fantastica idea della costruzione di una torre alta parecchi chilometri sulla quale porre un laboratorio spaziale. Dopo qualche anno l'idea della torre fu sostituita con quella di una lunghissima catena ancorata sulla Terra avente legato all'altra estremità un laboratorio. L'idea non aveva nulla di fantascientifico ma nessuno riteneva di potere disporre di un materiale capace di resistere alle enormi sollecitazioni cui sarebbe stato soggetto e fu presto abbandonata.

Non se ne parlò per molti anni ma nel 1972 Giuseppe Colombo, con la collaborazione del prof. Mario Grossi, ebbe l'idea di un laboratorio spaziale trattenuto da un filo non ancorato sulla Terra ma a una navicella spaziale, allo shuttle. In tal modo fu concepito il primo satellite a filo. Gli studi fatti dimostrarono che ciò è possibile utilizzando un filo che può arrivare sino a 100 chilometri e anche più.

L'idea ha un grande significato non solo per i risultati cui la sua attuazione può dar luogo ma anche perché, dopo la nascita dell'era spaziale avvcnuta con il lancio del primo sputnik, essa documenta quella dell'Uomo spaziale, dell'Uomo, cioé, capace di pensare come Uomo dello spazio. Solo questi infatti, non rigetta l'idea del satellite a filo, mentre chi pensa come Uomo della Terra è portato aprioristicamente a rifiutarlo, ritenendo istintivamente enorme e inaccettabile la tensione causata dalla gravità nel filo. Ma se si riesce a pensare come Uomo dello spazio, e Colombo fu certamente tale, si è portati ad approfondire la questione e si scopre che non è così: per trattenere un satellite il cui diametro non superi i due metri posto alla distanza di non oltre 50 chilometri dallo shuttle è sufficiente un filo il cui diametro è di soli due millimetri!

Il modello matematico del satellite a filo è quello della dinamica — o, in particolare, della statica — di un sistema formato da due corpi legati da un filo in presenza di forse gravitazionali, centrifughe, di Lorentz, elettromagnetiche, a variazioni di temperatura, ecc.. È un problema complesso che è stato studiato da Colombo con la collaborazione del

prof. Grossi, con particolare attenzione alle questioni di stabilità delle soluzioni. Ancor oggi vari ambienti scientifici si interessano a questioni aperte connesse a tale problema, in particolare ciò accade a Padova per opera di un gruppo di esperti formatisi alla scuola di G. Colombo.

Il progetto del satellite a filo è realizzabile con il satellite posto al di sopra della navicella di ancoraggio per la creazione di un osservatorio astronomico o con il satellite posto al di sotto per lo studio di questioni concernenti l'alta atmosfera: costituzione chimica, campo magnetico, gradiente di gravità, e così via. In biologia possono aversi interessanti utilizzazioni altrimenti impossibili; ad es., la realizzazione di un satellite con filo di ancoraggio di lunghezza variabile rende possibile la costituzione di un laboratorio a gravità variabile per lo studio del comportamento degli esseri viventi: nascita e crescita delle piante, ecc..

Ma la fantasia umana non ha limiti e alle ipotesi di utilizzazione or ora segnalate altre ne aggiunge, in parte indicate dallo stesso Colombo.

Esse possono sembrare fantastiche ma non hanno nulla di fantascientifico.

Sondaggi numerici mostrano la possibilità di ottenere enormi potenze nella produzione di corrente elettrica generata dal moto del filo nel campo magnetico terrestre; il fatto poi che il filo formerebbe circuito con il plasma circostante permetterebbe, per interazione con il campo magnetico presente, la creazione di un motore elettrico mediante l'invio di una corrente elettrica generata mediante lo sfruttamento di energia solare. Lo stesso Colombo ha studiato gli effetti dinamici dovuti a un improvviso taglio del filo: esso ne annullerebbe bruscamente la tensione causando un moto del satellite verso l'alto, se esso è posto al di sopra dello shuttle, e uno verso il basso della navicella di ancoraggio; ciò suggerirebbe un nuovo modo di rientro a casa dello Shuttle ma il problema è tutto da studiare.

L'idea del satellite a filo incontrò in un primo tempo molta diffidenza: l'Uomo dello spazio non esisteva ancora e sembrava inconcepibile che fosse possibile una tale struttura per l'incapacità di comprendere la diversità profonda che si ha ove la gravità è ridotta a ben poca cosa. Ma la personalità scientifica del proponente e le accurate e acute motivazioni portate a sostegno alla fine riuscirono ad avere il sopravvento e a vincere la diffidenza e le opposizioni altrui. Il progetto teorico divenne obiettivo concreto quando l'ente che aveva reso possibile lo sbarco dell'Uomo sulla Luna, il Marshall Space Flight Center, si interessò ad esso. La decisione di realizzarlo divenne definitiva quando essa fu suggellata da una lettera del Presidente Reagan al Presidente del Consiglio Italiano, nel 1983. Il lancio, previsto per il 1988-89 subirà qualche

ritardo anche a causa della nota recente catastrofe avutasi nell'attività spaziale americana.

Le conversazioni che io avevo con Giuseppe Colombo quando andavo a trovarlo nel Suo studio o quando Egli veniva nel mio vertevano spesso sul problema del satellite a filo e sulle complesse questioni matematiche ad esso collegate. Notavo l'enorme interesse del progetto, l'entusiasmo con cui Egli lo studiava ma anche — purtroppo — il Suo amaro convincimento, che una volta mi espresse esplicitamente, di non fare in tempo a vederne la realizzazione.

L'attività dello spirito costituisce una prerogativa unitaria dell'Uomo. Egli, spesso, nel penetrare i problemi della Natura con gli strumenti che la sua specifica preparazione gli fornisce, li sente, quasi inconsciamente con animo che è più dell'Artista o del Credente che dello Scienziato. È quanto accade in Giuseppe Colombo. Chi legge i Suoi scritti, specie quelli ove Egli espone in modo piano, rigoroso ma non tecnico, sintesi suggestive sulla conoscenza del mondo e ne rende partecipe i non esperti, vi trova un'intensa commozione che non è certo dello scienziato puro o del tecnico ma piuttosto di Chi penetra i misteri prima con l'animo che con la ragione, sentendone innanzitutto la suggestione. In un discorso del 1970, pubblicato sulla rivista «Cultura e Scuola» Colombo dichiara, testualmente: «Mentre il filosofo cerca dentro di se il senso del problema del mondo che ci circonda, l'uomo di scienza con un atto di fiducia illimitato continua a perseverare nel cammino della ricerca eccitato anziché depresso dal sempre più vasto, più profondo, più imprevedibile universo che si apre alla sua vista, come sempre nuovi cieli apparivano ai primi navigatori dei mari del sud. Nessuno di noi, testimoni di questa eccitante e travagliata epoca storica, qualunque sia il modo in cui ciascuno ha risolto il proprio problema spirituale di fronte al creato, è certo sfuggito al nodo della commozione, quando il primo uomo ha poggiato piede sulla superficie lunare portatovi da un'aquila prodotta dalla sua mente e dalle sue mani».

E in una relazione su Copernico, tenuta ai Lincei nel 1973, Colombo dichiara: «Che la rivoluzione Copernicana sia stata provocata da una crisi estetica più che da una crisi razionale, sembra abbastanza chiaro. La motivazione che doveva portare, nell'arco di un secolo, attraverso Keplero, Galilei e Newton, alla prima visione scientifica dell'universo, è di natura squisitamente estetica più che razionale. È una tendenza, forse inconscia, alla ricerca appassionata di simmetrie, di bene ordinati sistemi, di strutture geometriche regolari.» e poi, riferendosi a Copernico, aggiunge: «... Impone idee sconvolgenti e rifiutate perché contrarie ad

un *senso comune* tramandato da millenni. E tutto ciò per una esigenza profonda di trovare quell'ordine nell'Universo che solo una sensazione interiore gli fa percepire.»

In un altro scritto, Colombo, richiamando la preghiera con la quale Keplero chiude l'Harmonices Mundi trova in essa «l'intuizione, anzi la certezza, che tra idee fantasiose c'erano anche i motivi conduttori della sinfonia del creato».

Giuseppe Colombo che tanto intensamente ha rivolto lo sguardo verso il cielo, fu un uomo legato soprattutto alla Sua Terra, al nostro pianeta che Egli ha tanto amato e studiato ma che ha dovuto lasciare prematuramente. Nel discorso su Copernico da me prima richiamato, Egli definisce la Terra «il nostro pianeta pieno di colori, l'unico disponibile per noi» e afferma: «Risolvere il problema della storia della Terra è una necessaria premessa per risolvere il problema della vita, il problema delle origini dell'uomo. Perché, come dice Shapley, noi siamo fratelli delle pietre e eugini delle nuvole.»

A tre anni dalla morte di Giuseppe Colombo ognuno di noi porta vivo il ricordo della Sua figura, accostante ed amichevole. Ci incontravamo spesso, specialmente negli ultimi tempi della Sua vita terrena, quando Egli non andava frequentemente in America come prima. Si parlava di tante cose e di tante questioni di comune interesse scientifico, come se nulla di terribile dovesse accadere. Ma Egli ci ha lasciato prematuramente, allontanandosi anzitempo dai Suoi cari, dai Suoi collaboratori ed allievi, dai Suoi amici. A tutti è rimasta in eredità la Sua opera profonda di studioso del nostro sistema solare.

Egli ha lavorato il più a lungo possibile, sino all'ultimo con apparente indifferenza e senza cercare di coinvolgere i suoi interlocutori. Un modesto sfogo lo ebbe solo quando, invitato a tenere il discorso di chiusura dell'anno accademico dell'Istituto Veneto, iniziò con queste parole: «Fino a poco tempo fa non ero sicuro che avrei avuto questa opportunità di incontrare tanti cari amici che mi hanno seguito in passato con tanto affetto. In questo straordinario ambiente, questa cerimonia di chiusura dell'Anno Accademico dell'Istituto Veneto mi dà l'idea di una giubilazione personale.

Infatti, l'invito rivoltomi dai colleghi in tempi per me molto oscuri mi ha lasciato perplesso, per dir poco. Ho comunque preso il coraggio a due mani ed ho cominciato a pensare come impostare un discorso che si addicesse alla solennità della manifestazione».

Con queste parole Giuseppe Colombo ci ha dato una lezione di vita, mostrandoci il Suo amore per la scienza da Lui prediletta.

BIBLIOGRAFIA
Giuseppe Colombo

1952 *Sopra un sistema non-lineare in due gradi di libertà.* Rend. Sem. Mat. Padova, vol. 21, pp. 64-98.

1953 *Sul moto di due corpi rigidi pesanti collegati in un punto, di cui uno ha un punto fisso.* Rend. Sem. Mat. Padova, vol. 22, pp. 305-312.

1953 *Sopra una questione di ottica geometrica.* Rend. Accad. Lincei, ser. VIII, vol. 14, pp. 627-631.

1954 *Limitazioni superiori per i moduli delle componenti di stress in un particolare problema di deformazione piana.* Ann. Univ. Ferrara, ser. VII, vol. 3, pp. 45-54.

1954 *Sui sistemi autonomi di ordine superiore al secondo.* Istituto Matematico, Roma 25 pp.

1954 *Sulle oscillazioni forzate di un circuito comprendente una bobina a nucleo di ferro.* Rend. Sem. Mat. Padova, vol. 23, pp. 407-421.

1955 *Oscillazioni persistenti di un sistema non lineare dissipativo dovute al ritardo della forza di richiamo.* Ann. Univ. Ferrara ser. VII, vol. 4, pp. 33-50.

1955 *Maggiorazioni delle componenti di stress nel problema di De Saint-Venant.* Rend. Sem. Mat. Padova, vol. 24, pp. 70-83.

1955 *Sopra un problema della dinamica del binario.* Rend. Sem. Mat. Padova, vol. 24, pp. 230-244.

1955 *Sopra il fenomeno dell'azione asincrona.* Rend. Sem. Mat. Padova, vol. 24, pp. 353-395.

1955 *Moti di un sistema non-lineare autonomo in due gradi di libertà, con debole accoppiamento capacitivo.* Rend. Sem. Mat. Padova, vol. 24, pp. 400-420.

1955 *Riduzioni alle quadrature di un notevole problema di stereodinamica.* Rend. Accad. Lincei, ser. VIII, vol. 18, pp. 168-172.

1957 *Sopra un notevole fenomeno nel campo delle vibrazioni non lineari di combinazione.* Rend. Accad. Lincei, ser. VIII, vol. 22, p. 726-730.

1957 *Sulle oscillazioni non-lineari di combinazione.* Rend. Sem. Mat. Padova, vol. 27, pp. 162-175.

1958 *Sopra il problema del «Lacet»*. Journ. Math. Mech., vol. 7, pp. 483-502.

1958 *Teoria del regolatore di Bouasse e Sarda*. Rend. Sem. Mat. Padova, vol. 28, pp. 338-347.

1958 *Sulle orbite stabili in un sincrotrone con tratti rettilinei*. Ann. Mat. Pura Appl., serv. IV, vol. 46, pp. 249-263.

1961 *The motion of satellite 1958 Epsilon around its center of mass*. Smithsonian Astrophys. Obs. Spec. Rep. No. 70, 26 pp.

1961 *The stabilization of an artificial satellite at the inferior conjunction point of the earth-moon system*. Smithsonian Astrophys. Obs. Spec. Rep. No. 80, 13 pp.

1962 *Instability of motion at the Lagrangian triangular point in the earth-moon system*. Nature, vol. 193, p. 1063.

1962 *On the motion of Explorer XI around its center of mass. Smithsonian Astrophys.* Obs. Spec. Rep. No. 94, 25 pp.

1962 *On some singular orbits of an earth-moon satellite with a high area-mass ratio (abstract) (G. Colombo and D.A. Lautman)*. Astron. Journ., vol. 67, p. 573; also in Smithsonian Astrophys. Obs. Spec. Rep. No. 107, 14 pp.

1963 *On the libration orbits of a particle near the triangular point in the semi-restricted three-body problem (abstract) (G. Colombo, D.A. Lautman and C.M. Munford)*. Astron. Journ., vol. 68, pp. 159-162; also in Smithsonian Astrophys. Obs. Spec. Rep. No. 108, 10 pp.

1963 *Optical radar results and meteoric fragmentation (G. Colombo and G. Fiocco)*. Smithsonian Astrophys. Obs. Spec. Rep. No. 139, 25 pp.

1963 *The magnetic torque acting on artificial satellites*. Proc. Conf. on Gyrodynamics, IUTAM, Celerina. In *Kreisel Probleme Gyrodynamics* ed. by H. Ziegler, Springer-Verlag, Berlin.

1963 *Capture of cosmic dust into circumterrestrial orbits (abstract). (G. Colombo, I.I. Shapiro, and D.A. Lautman)*. Trans. Amer. Geophys. Union, vol. 44, p. 71.

1964 *The capture of cosmic dust by the earth* (G. Colombo, D.A. Lautman and 1.1. Shapiro). Presented at the 5th International COSPAR Meeting Florence, Italy, May.

1964 *Optical radar results and meteoric fragmentation (with G. Fiocco)*. Journ. Geophys. Res., vol. 69, no. 9, May.

1964 *Thermal effect on the rotational period of an artificial satellite* (G. Colombo and P. Higbie). Smithsonian Astrophys. Obs. Spec. Rep. No. 166, 22 pp.

1965 *Rotation period of the planet Mercury.* Nature, vol. 208, p. 575.

1965 *Reply to «note on laser detection of atmospheric dust layers» by D. Deirmendjian (with G. Fiocco).* Journ. Geophys. Res., Vol. 70, p. 746.

1965 *The rotation of the planet Mercury* (G. Colombo and I.I. Shapiro). Smithsonian Astrophys. Obs. Spec. Rep. No. 188R, 24 pp.

1965 *Dynamics of small particles in the solar system* (G. Colombo, D.A. Lautman and I.I. Shapiro). Presented at the Annual Meeting of the American Geophysical Union, Washington, April.

1965 *Cassini's second and third laws.* Presented at the International Conference on Selenodesy, Manchester, England, May; also in Smithsonian Astrophys. Obs. Spec. Rep. No. 203, 19 pp.

1965 *The earth's dust belt: Fact of fiction?* (G. Colombo, I.I. Shapiro and D.A. Lautman). Presented at the Symposium on Meteoritic Orbits and Dust, Smithsonian Astrophys. Obs., Cambridge, Massachusetts, August.

1965 *Recent developments in theory of rotation of Mercury and Venus.* Presented at the Meeting on Mantle of the Earth and Terrestrial Planets, NATO Advanced Study Institute, Newcastle-upon-Tyne, April.

1967 *Theory of the axial rotations of Mercury and Venus* (with E. Bellomo and I.I. Shapiro). in *Mantles of the Earth and Terrestrial Planets*, Interscience Publ., New York, pp. 194-211.

1968 *On a family of periodic orbits of the restricted three body problem and the question of the gaps in the asteroid belt and in Saturn's rings* (with F.A. Franklin and C. Munford). Astron. Journ., vol. 73, pp. 111-123.

1969 *A theory of the dynamical structure of the Saturn ring supported by recent observations.* Mem. Soc. Astron. ital., vol. XL, fasc. 2, pp. 205-209.

1970 *A dynamical model for the radial structure of Saturn's rings* (F.A. Franklin and G. Colombo). Icarus, vol. 12, pp. 338-347.

1970 *A model of radial structure of the rings of Saturn* (G. Colombo and F.A. Franklin). In *Symposia Mathematica*, vol. III, Academic Press, London, pp. 65-74.

1971 *A dynamical model for the radial structure of Saturn's rings.* II (F.A. Franklin, G. Colombo, and A.F. Cook). Icarus, vol. I5, pp. 80-92.

1971 *On the formation of the outer satellite groups of Jupiter* (G. Colombo and F.A. Franklin). Icarus, vol. I5, pp. 186-I89.

1974 *On the formation of the orbit-orbit resonance of Titan and Hyperion* (G. Colombo, F.A. Franklin, and I.I. Shapiro). Astron. Journ., vol. 79, pp. 61-72.

1974 *The role of the asteroid belt in the early history of the inner solar system* (G. Colombo and F.A. Franklin). Presented at the International Meeting on Physics and Geology of Terrestrial Planets, Accademia Nazionale dei Lincei, Rome, April.

1974 *Shuttle-borne «Skyhook»: A mew tool for low-orbital-altitude research* (G. Colombo, E.M. Gaposchkin, M.D. Grossi and G.C. Weiffenbach). Smithsonian Astrophys. Obs. Reports in Geoastronomy No. 1, 64 pp.

1975 *Long-tethered satellites for the Shuttle Orbiter* (G. Colombo, E.M. Gaposchkin, M.D. Grossi and G.C. Weiffenbach). Presented at the International Conference on Technology of Scientific Space Experiments, Paris, France, May.

1975 *The «Skyhook»: A Shuttle-borne tool for low-orbital-altitude research* (G. Colombo, E.M. Gaposchkin, M.D. Grossi and G.C. Weiffenbach). Meccanica, no. 1, vol. X, pp. 3-20.

1975 *Shuttle-borne long-tethered satellites: A new tool for low-cost space sciences and applications* (G. Colombo, E.M. Gaposchkin, M.D. Grossi and G.C. Weiffenbach). Presented at the American Astronomical Society/American Institute of Aeronautics and Astronautics Astrodynamics Conference, Nassau, Bahamas, July.

1976 *An alternative option to the dual-probe out-of-ecliptic mission via Jupiter swing-by* (G. Colombo, D.A. Lautman and G. Pettengill). In *Proceedings of the Symposium on the Study of the Sun and Interplanetary Medium in Three Dimension*, ed. by L.A. Fisk and W.I. Axford, NASA TM-X-71097. pp- 37-47.

1976 *Dumbbell gravity-gradient sensor: A new application of orbiting long tethers* (G. Colombo, D.A. Arnold, J.H. Binsack, R.H. Gay, M.D. Grossi, D.A. Lautman and O. Orringer).
Smithsonian Astrophys. Obs. Reports in Geoastronomy No. 2, 59 pp.

1976 *Secular resonance, solar spin down and the orbit of Mercury* (W.R. Ward, G. Colombo and F.A. Franklin). Icarus, vol. 28, pp. 441-452.

1976 *Electrodynamics of long tethers in the near-earth environment* (M. Dobrowolny, G. Colombo and M.D. Grossi). Final Rep., NASA Contract NAS8-31678, June; also in Smithsonian Astrophys. Obs. Reports in Geoastronomy No. 3, 52 pp.

1976 *An arrow to the sun* (J.D. Anderson, G. Colombo, L.D. Friedman and E.L. Lau). Presented at the International Symposium on Experimental Gravitation, Pavia, Italy, September; proceedings in press.

1976 *Gravity-gradient measurements down to 100-km height by means of long-tethered satellites* (G. Colombo, E.M. Gaposchkin, M.D. Grossi and G.C. Weiffenbach). Presented at the 27th Conference of the International Astronautical Federation, Anaheim, California, October.

1976 *Atmospheric modifications by a high-power microwave beam at 22.2 GHz (water-vapor absorption line)* (M.D. Grossi and G. Colombo). Presented at the 1976 International IEEE/AP-S Symposium and USNC/URSI Meeting, Amherst, MA, October.

1976 *Spiral structure as an explanation for the asymmetric brightness of Saturn's A ring* (letter) (G. Colombo, P. Goldreich, and A.W. Harris). Nature, vol. 264, pp. 344-345.

1977 *An arrow to the sun* (J.D. Anderson, G. Colombo, L.D. Friedman, and E.L. Lau). In *Proceedings of the International Meeting on Experimental Gravitation*, Roma, Accademia dei Lincei, pp. 393-422.

1977 *Spaceborne high-power microwave beam at 22.2 GHZ (water vapor absorption line) and its potential for atmospheric modifications* (M.D. Grossi and G. Colombo). Journ. Microwave Power, vol. 12, pp. 209-213.

1977 *Tethered satellite system for gravity gradiometry at low orbital height* (G. Colombo, D.A. Arnold, E.M. Gaposchkin, M.D. Grossi, P.M. Kalaghan, and L.R. Kirschner). Presented at the Earth Dynamics Summer Workshop, Boulder, Colorado, July.

1977 *An alternative option to the dual-proble out-of-ecliptic mission via Jupiter swing-by* (G. Colombo, D. Lautman, and G. Pettengill). Presented at the International Conference «M. Panetti», Torino, Fall.

1977 *Advanced technologies in space and opportunities for gravity experiments* (C.A. Lundquist and G. Colombo). Presented at the International School of Relativistic Astrophysics, Trapani, Italy, Fall.

1978 *On the azimuthal brightness variations of Saturn's rings* (F.A. Franklin and G. Colombo). Icarus, vol. 33, pp. 279-287.

1978 *Study of the dynamics of a tethered satellite system* (SKYHOOK) (P.M. kalaghan, D.A. Arnold, G. Colombo, M.D. Grossi, L.R. Kirschner, and O. Orringer). Final Rep., NASA Contract NAS8-32199, March.

1978 *Gravity gradient determination with tethered systems* (P.M. Kalaghan and G. Colombo). Presented at the NASA/MSFC Workshop on the Uses of a Tethered Satellite System, Huntsville, Alabama, May.

1978 *Interactions of a tethered satellite system with the ionosphere* (M.D. Grossi and G. Colombo). Presented at the NASA/MSFC Workshop on the Uses of a Tethered Satellite System, Huntsville, Alabama, May.

1978 *The role of the asteroid belt in the early history of the inner solar system* (G. Colombo and F.A. Franklin). In *Proceedings of the International Meeting onPhysics and Geology of Terrestrial Planets.* In press?

1979 *Mechanisms of electrodynamic interactions with a tethered satellite system and the ionosphere* (M. Dobrowolny, D.A. Arnold, G. Colombo, M.D. Grossi). Reports in Radio and Geoastronomu No. 6, August.

1979 *System noise analysis of the Dumbbell tethered satellitge for gravity-gradient measurements* (D.A. Arnold, G. Colombo, N. Lanham, and G. Nystrom). Final Report NASA Grant NSG-8063.

1979 *Space science and application program of the 80's and beyond: An overview of future user's requirements.* European Space Agency, DP/ST (79) 4.

1979 *A preliminary study of the orbiting platform-tether system space operation center. Department of Aeronautics and Astronautica, MIT.*

1980 *Mathematics, sciences and mathematical sciences.* Pontificiae Academiae Scientiarum Commentarii.

1981 *Tether system safety study: Preliminary analysis.* Progress report NASA Contract NAS8-33691.

1981 *Study of certain launching techniques using long orbiting tethers.* Final Technical Report, NASA Grant 8008.

1981 *Investigation of electrodynamic stabilization and control; Report on Contract NAS8-33691,* by G. Colombo (PI), M.D. Grossi, M. Dobrowolny, and D.A. Arnold, March.

1981 *Satellite-to-satellite doppler tracking (SSDT) for mapping of the Earth's gravity field, Final Report on NAG5-36,* by G. Colombo, E.M. Gaposchkin, M.D. Grossi and R.D. Estes, September.

1982 *The use of tethers for payload orbital transfer, Report on Contract NAS-33691,* by G. Colombo (PI), M. Martinez-Sanchez and D.A. Arnold, March.

1982 *A possible link between the rotation of Saturn and its ring structure* (with F.A. Franklin and A. Cook) Nature, vol. 295, January.

1982 *Frome the «Dialog of Galilei to the Modern Ones. Invited paper delivered in Rome at the Academy of Lincei on occasion of the celebration of the 350th anniversary of the publication of the «Dialogo sui due massime sistemi del mondo».* May 1982.

1982 *Theory of motion of Saturn's co-orbiting satellites: 1980 S1 and S3* (with C.F. Yoder and S.F. Synnot) submitted to Icarus in Spring 1982 (in revision)?

1982 *The Mariner 10 MUltiple Fly-by to Mercury: Two accurate fixes of Mercury* (with J.D. Anderson and E. Lau) to be submitted to the Astronomical Journal.

1982 *One new concept for a Space Station Architecture* (with J.W. Slowey). Published with the Report of the Workshop on the Utilization of the External Tanks of the Space Transportation System held at the Space Institute/USCD, Agust 23-27.

1982 *An Alternative Configuration of the Electrodynamic Tether Experiment*, Technical Note TP 82-001, February 8.

1982 *Use of Shuttle External Tanks in a Large Space Platform*, Technical Note TP 82-002, March 15.

1982 *A Straightforward Use of the Shuttle External Tank in a Large Space Platform*, Technical Note TP 82-003, April 14.

1982 *On a New Concept for a Space Station Architecture*, Technical Note TP 82-005 (with J.W. Slowey), September 16.

1983 *Orbital Transfer and Release of Tethered Payloads* (with M.D. Grossi, M. Martinez-Sanchez and D. Arnold) Final Report NAS8-33691, March.

1983 *Dynamic Noise in Tethered Subsatellite: Requirement for Gyro Angular Rotationl Stability*, Technical Note TP 83-004, May 23.

New generalizations of the concept of the gravitational restricted problem of many bodies

Victor G. SZEBEHELY *

Abstract. *The gravitational problem of three bodies was first treated from the point of view of dynamics by Sir Isaac Newton who obtained some classical results in the lunar theory (1687). The simplification known as the restricted problem of three bodies was introduced by L. Euler in his lunar theory (1772). Both the general and the restricted problems of three bodies are non-integrable as shown by H. Poincaré (1890). The degrees of freedom in the general formulation are 9, while in the restricted formulation is only 3. As a consequence, the restricted problem, whenever applicable, became one of the cornerstones of celestial mechanics. With the space-age came problems of trajectories of space vehicles which are the best examples satisfying the requirements of the restricted problem. The simplifications offered by the restricted formulation are applicable to a great variety of dynamical problems and a systematic description of such systems is the subject of this paper. As new application, the dynamics of gravitationally interacting minor bodies in the solar system are formulated and the dynamics of binary asteroids are discussed.*

A new definition of the restricted problem

The basic idea behind the dynamical model know as the restricted problem is that when the magnitudes of the participating masses are ordered, the motion of those bodies which have much larger masses than others might be considered uneffected gravitationally by the bodies of much smaller masses.

This concept is generally accepted in the literature, with few exceptions. These exceptions restrict the definition of the restricted problem to the motion of three point masses, two of which have much larger masses than the third. The two large masses move on circular orbits around their center of mass and are not influenced by the third body. The problem consists of determining the behavior of the body with the small mass. This classical restricted problem has many generalizations, such

* University of Texas, Austin, Tx.

as the elliptic restricted problem (when the large masses move on elliptical orbits), the variable mass restricted problem, the perturbed restricted problem (for instance, considering solar radiation pressure on an Earth-Moon-space probe problem), the finite mass (as opposed to point mass) restricted problem, etc. If neither the number of bodies, nor the various details of the formulation of the system are specified and only the basic principle, as described in the introductory sentence of this chapter is operational, the general concept of the restricted problem emerges. In this paper, this definition is accepted for the restricted problem and, therefore, the classical restricted problem of three bodies will be treated as one of the simplest special cases.

Returning now to the general definition we note that the bodies with larger masses will effect each other as well as will influence the motion of the bodies with smaller masses. The smaller bodies might effect the motion of each other or might not and we speak about gravitationally interacting or non-interacting bodies but they will not influence the motion of the large bodies.

In this way a unified concept of the restricted problem is arrived at and this will allow us the presentation of a hierarchy of restricted problems.

The hierarchical system of restricted problems

The hierarchical restricted problem consists of N systems each having several bodies. The first system consists of n_1 bodies with masses m_1^1, $m_2^1, \ldots, m_{n_1}^1$. The bodies of this system will influence the motion of each other and also of all other bodies belonging to all the other systems. The second system consists of n_2 bodies with masses $m_1^2, m_2^2, \ldots, m_{n_2}^2$. These bodies will *not* influence the motion of the bodies in the first system but will influence each other and all bodies in the other systems. The bodies of the first system have larger masses than the bodies belonging to the second system, or $m_\ell^1 \gg m_j^2$, where $1 \leqslant \ell \leqslant n_1$, and $1 \leqslant j \leqslant n_2$. The third system consists of n_3 bodies with masses m_1^3, $m_2^3, \ldots, m_{n_3}^3$. These bodies are influenced by each other and by the bodies of the first and second systems. The bodies belonging to this (third) system do not influence the motion of the bodies of systems 1 and 2 but they influence the motion of bodies belonging to systems 3, 4, ... N. In general, system i consists of masses m_1^i, $m_2^i, \ldots m_{n_i}^i$, satisfying the inequalities $m_p^q \gg m_\ell^i \gg m_k^j$, where $1 \leqslant p \leqslant n_q$, $1 \leqslant \ell \leqslant n_i$, $1 \leqslant k \leqslant n_j$, $q < i < j$.

Note that superscripts show the system in which the mass is located

and subscripts identify the particular body. System i influences the motion of masses belonging to systems $j \geqslant i$, that is bodies in system i and bodies belonging to systems $i+1$, $i+2, \ldots N$.

Bodies not influenced by system i belong to systems $q < i$, that is to systems $i-1$, $i-2, \ldots 2$, 1.

The last system is denoted by N. It contains n_N bodies with masses m_1^N, $m_2^N, \ldots m_{n_N}^N$ which are the bodies with the smallest masses in the hierarchy. These bodies influence the behavior of each other, are influenced by all other bodies belonging to all other systems and do not influence the motion of bodies of any of the other systems.

The total number of systems is N, the total number of bodies is given by:

$$\sum_{i=1}^{N} n_i.$$

The i-th body in the last system has mass m_i^N and its position vector is given by $\bar{r}_i^N$. The equation of motion of this body might be written as:

$$\ddot{\bar{r}}_i^N = -G \sum_{\alpha=1}^{N} \sum_{j=1}^{n_\alpha} \frac{\bar{r}_i^N - \bar{r}_j^\alpha}{|\bar{r}_i^N - \bar{r}_j^\alpha|^3} m_j^\alpha,$$

where, when $\alpha = N$, $i \neq j$. This equation might be derived by considering that all bodies will influence the motion of m_i^N. The term in the summation corresponding to $\alpha = N$ and $i = j$ is to be eliminated since it would correspond to the influence of the mass m_i^N on itself.

The internal summation represents the forces acting on mass m_i^N by bodies of system α. The second (outside) summation corresponds to including all systems from the largest masses ($\alpha = 1$) to the smallest masses ($\alpha = N$). If the smallest masses do not interact gravitationally then the outher summation is to be performed from $\alpha = 1$ to $\alpha = N-1$.

Special cases

In the classical restricted problem of three bodies we have the system of the primaries, with masses m_1^1 and m_2^1 and the next system with only one small mass $m_1^2 \ll m_i^1$, where $1 \leqslant i \leqslant 2$. If there are two small masses belonging to system 2, we have $m_i^1 \gg m_j^2$, where $i = 1$, 2 and

$j = 1, 2$. The problem of binary asteroids belongs to this category. The bodies with masses m_1^1 and m_2^1 are the Sun and Jupiter and the masses of the two asteroids forming the binary are m_1^2 and m_2^2. The relation between the masses becomes

$$m_1^1 \geq m_2^1 \gg m_1^2 \geq m_2^2.$$

The lunar problem might be approximated by $m_1^1 = m_\odot$, $m_2^1 = m_\oplus$, $m_1^2 = m_☾$.

Satellites of planets in a stellar system might be represented by three systems. First the masses of the stars $m_1^1, m_2^1, \ldots m_{n_1}^1$, then the planets $m_1^2, m_2^2, \ldots m_{n_2}^2$ and finally the masses of the satellites $m_1^3, m_2^3, \ldots m_{n_3}^3$. The planets will not influence the motion of the stars and the satellites will not influence the motion of the stars and the satellites will not influence the motion of the planets in general. If space probes are present in the system, we have $m_1^4, m_2^4, \ldots m_{n_y}^4$. In general

$$m_i^1 \gg m_j^2 \gg m_k^3 \gg m_\ell^4 .$$

The stars in this system will interact with each other and they will influence the motion of the planets, of the satellites and of the space probes. The planets might interact with each other or might not and will influence satellites, especially their own satellites and the space probes. The satellites might or might not interact with each other and will influence nearby space probes. The space probes, in general will not influence the motion of any of the bodies but could be affected by all bodies, depending on their orbits.

Conditions for application of the restricted problem

The basic principle of applicability might be established considering two systems of the hierarchy, the system with the largest masses, m_1^1, $m_2^1, \ldots m_{n_1}^1$ and the next system, $m_1^2, m_2^2, \ldots m_{n_2}^2$. The condition of applicability of the principle of the restricted problem in this case is that the bodies with masses m_j^2 do not influence the motion of bodies with masses m_i^1. As we will see, the basic idea might be demonstrated with having two bodies in the first system, with masses m_1^1 and m_2^1 and one body in the second system with mass m_1^2. The equations of motions of these three bodies for the general case might be written as follows:

$$\ddot{\bar{r}}_1^1 = -G\left[m_2^1 \frac{\bar{r}_1^1 - \bar{r}_2^1}{|\bar{r}_1^1 - \bar{r}_2^1|^3} + m_1^2 \frac{\bar{r}_1^1 - \bar{r}_1^2}{|\bar{r}_1^1 - \bar{r}_1^2|^3}\right]$$

$$\ddot{\bar{r}}_2^1 = -G\left[m_1^2 \frac{\bar{r}_2^1 - \bar{r}_1^2}{|\bar{r}_2^1 - \bar{r}_1^2|^3} + m_1^1 \frac{\bar{r}_2^1 - \bar{r}_1^1}{|\bar{r}_2^1 - \bar{r}_1^1|^3}\right]$$

$$\ddot{\bar{r}}_1^2 = -G\left[m_1^1 \frac{\bar{r}_1^2 - \bar{r}_1^1}{|\bar{r}_1^2 - \bar{r}_1^1|^3} + m_2^1 \frac{\bar{r}_1^2 - \bar{r}_2^1}{|\bar{r}_1^2 - \bar{r}_2^1|^3}\right]$$

The principle of the restricted problem was not applied here. The first equation representing the motion of m_1^1 shows the influence of m_2^1 as well as of m_1^2 on m_1^1. The first term of the second equation shows the effect of m_1^2 on m_2^1 and the second term represents the effect of m_1^1 on m_2^1. The third equation shows the effect of m_1^1 and m_2^1 on m_1^2. The assumption considering the magnitude of masses is that the masses of the first system are much larger than the mass of the body belonging to the second system, or

$$m_1^1,\ m_2^1 \gg m_1^2.$$

Considering only the scalar magnitudes of the terms in the equations of motion and introducing a simplified notation, we might write the six terms in the following table:

$$\frac{m_2}{\Delta_{12}^2} \qquad \frac{m_3}{\Delta_{31}^2}$$

$$\frac{m_3}{\Delta_{23}^2} \qquad \frac{m_1}{\Delta_{12}^2}$$

$$\frac{m_1}{\Delta_{31}^2} \qquad \frac{m_2}{\Delta_{23}^2}$$

The new symbols are related to the previous (general) notations by the equations:

$$m_1 = m_1^1, \quad m_2 = m_2^1, \quad m_3 = m_1^2,$$

$$\Delta_{12} = |\bar{r}_1^1 - \bar{r}_2^1|, \quad \Delta_{23} = |\bar{r}_2^1 - \bar{r}_1^2|, \quad \Delta_{31} = |\bar{r}_1^2 - \bar{r}_1^1|.$$

The third equation of motion refers to the body with the small mass which is affected by the bodies with the large masses. Correspondingly, the last two terms in the table are not involved in the process of formulating the restricted problem. The first two terms of the table represent the effects of the large and of the small mass on the first large mass m_1. The condition of applicability of the restricted problem is that the second term might be neglected when compared to the first term, or

$$1 \gg \frac{m_3}{m_2} \left(\frac{\Delta_{12}}{\Delta_{31}} \right)^2 .$$

Similarly, the second line of the table represents the effects of the small mass (m_3) and the first large mass (m_1) on the second large mass (m_2). Once again, the condition of applicability of the restricted problem might be formulated by the requirement:

$$1 \gg \frac{m_3}{m_1} \left(\frac{\Delta_{12}}{\Delta_{23}} \right)^2 .$$

The above two inequalities demonstrate the complicated nature of the conditions of applicability, since not only the masses of the participating bodies enter but also the varying relative distances appear in the requirements. This is not surprising since the Newtoniana formulation of gravitation forces include masses *and* distances. The inequalities may be satisfied, however, as we will see, unless $\Delta_{12} \to \infty$. This means that the system of the bodies with the largest masses breaks up. This possibility however, might be excluded by reformulating the problem after break-up. Then the first inequality requires the investigation of the case when $\Delta_{31} \to 0$ and the second the $\Delta_{23} \to 0$ case. The smallest value Δ_{31} might take is the sum of the radii of the bodies m_1 and m_3, which might be denoted by $R_1 + R_3 \cong R_1$. Similarily for the second inequality we have $\mathrm{Min}\, \Delta_{23} = R_2 + R_3 \cong R_2$. To further estimate the requirements of the inequalities let us assume that the densities of the participating spherical bodies are about the same and that the radii of the bodies with the large masses are of the same order of magnitude, $R_1 \cong R_2 \cong R$. In this case

the two inequalities become identical and might be written as:

$$1 \gg \frac{R_3^3 \Delta_{12}^2}{R^5}.$$

Note that in the above simplification, the smallest values of Δ_{23} and Δ_{3}, correspond to collisions.

The evaluation of the above inequality for special cases requires the knowledge of the largest distance between the large masses (Max Δ_{12}), the smallest radius of the large bodies (R) and the radius of the body with the small mass (R_3).

For satellite or planetary problems when the Δ_{ij} quantities are approximately constants, the evaluation of the two original inequalities is simple. For instance for the Sun-Earth-Moon case, the first inequality becomes:

$$1 \gg \frac{m}{m_\oplus}\left(\frac{\Delta_{\odot\oplus}}{\Delta_{\oplus☾}}\right)^2 \cong 0.01$$

and the second:

$$1 \gg \frac{m_☾}{m_\odot}\left(\frac{\Delta_{\odot\oplus}}{\Delta_{\oplus☾}}\right)^2 \cong 0.006.$$

The meaning of the first inequality is that the Earth's effect on the Sun is approximately 100 times as large as the Moon's effect. Neglecting the Moon's effect and using the model of the restricted problem results in an error of 1% when the equation of motion of the Sun is formulated. The second inequality shows that the effect of the Sun on the Earth is approximately 173 times as large as the Moon's effect. Neglecting the Moon's effect results in an error of 0.6%.

The other satellites of the solar system satisfy the requirements of the inequalities with even higher accuracy than the Earth's Moon. An exception is Neptune's satellite Triton as can be seen using the above inequalities with $m_1 = m_\odot$, $m_2 = m_♆$, m_3 for the mass of Triton. The complete and systematic evaluation of the Solar System concerning applicability of the concept of the restricted problem will be presented elsewhere.

REFERENCES

EULER L., «Theoria Motuum Lunae», Acad. Imp. Scient., Petropoli, 1772.

NEWTON I., «Philosophiae Naturalis Principia Mathematica», Royal Society, London, 1687.

POINCARÉ H., «Les Méthodes Nouvelles de la Mécanique Céleste», Gauthier-Villars, Paris, 1892-1899.

SZEBEHELY V., «Theory of Orbits», Academic Press, New York, 1967.

Hierarchical dynamical n-Body systems and the stability of the solar system

Archie E. ROY *

Abstract. *The Solar System is an example of a hierarchical n-body dynamical system (n-body HDS); the problems of its dynamical age and long-term stability are considered, using methods formulated recently such as the general 3-body energy-angular momentum stability criterion, long-term numerical integrations, the empirical stability parameter method, the mirror theorem and its application to the stability of a system of bodies in almost circular almost coplanar orbits.*

Introduction

The maior bodies in the Solar System planetary and satellite systems are found in almost circular, almost coplanar, well spread out orbits which over many periods of revolution show little change in their semi-major axes, eccentricities and inclinations. The observed changes in these elements seem to be purely periodic, no secular trend having been observed. Changes in the other elements, though secular, have no effect on the stability of such systems.

Such systems are examples of hierarchical dynamical n-body systems (n-body HDS's). It is possible to order the bodies in an HDS by size of orbit with no crossing of orbits; it is usual in a simple HDS to take each member of the HDS to be in a perturbed Keplerian elliptic orbit about the centre of mass of those bodies moving in smaller orbits. The well-known Jacobi coordinate system is designed to express this ordering in the resulting equations of motion. In a complex HDS, of which the simple HDS is a special case, the hierarchy is generalized to take into account the more general possibility of the centre of mass of a sub-group of bodies perfoming a perturbed Keplerian orbit about the centre of mass of another sub-group of bodies in the HDS. Recent work on such complex HDS's may be found in Walker 1983, Walker and Roy 1983b, Milani and Nobili 1983 a, b.

* Department of Physics and Astronomy, Glasgow University, Glasgow, U.K.

There are many definitions of the stability of a dynamical system; in this paper the definition used is one of *hierarchical stability*. The HDS is said to exhibit hierarchical stability if the ordering of the hierarchy does not change in a time interval long compared with the longest period of revolution in the system, that is, no crossing of orbits or the escape of bodies from the system takes place within that time. There is nothing absolute about such a definition and in fact only the general three-body problem can have a hierarchy where such stability may be shown to hold for all time given certain initial conditions.

The problem of the age and long-term stability of the Solar System is one that has exercised the minds of celestial mechanicians for over three centuries. As far as the ages of the bodies are concerned, geophysics, lunar sample and meteorite dating and solar astrophysics suggest agreeably close answers in the range 4.5 to 5×10^9 years. The fossil record shows that complicated life forms have inhabited the Earth for some 2×10^9 years so that during that time the Sun's radiation output cannot have altered to any major extent; nor can the major or minor distances of Earth from Sun have strayed far from their present values. Even today, however, celestial mechanics is still unable to make confident statements on the dynamical age, stability and evolution of the Solar System though much progress has undoubtedly been made in recent years in many parts of the general problem, such as the restricted and general three-body problem, the stability of planetary rings, empirical stability studies of HDS's and so on.

Analytical theories and numerical integrations

First formulated by Newton, the point-mass Newtonian gravitational n-body problem equations of motion admit ten integrals cnabling several general and useful statements to be made. Known to Euler, they say that the total energy H and angular momentum c of the system are constant and that the centre of mass travels with constant velocity. No further general integrals have been found though special solutions of the three-body problem were found by Lagrange including the equilateral triangular solution, now exemplified not only by Trojan asteroids but also by the companions of Dione and Tethys in Saturn's satellite system. Lagrange, Laplace, Poisson and Leverrier, among others, produced planetary perturbation theories in which analytical expressions for the orbital elements, in essence long series of sines, cosines and secular terms, give highly accurate values of the orbital elements and positions

of the planets for time intervals of thousands of years past or future. Such expressions, however, were shown by Poincaré in 1895 to be in general divergent and inapplicable to questions of the long-term stability of the Solar System. More recent work culminating in the well-known Kolmogorov-Arnold-Moser theorem (the KAM theorem) has shown that within certain parameter value ranges, series expression are convergent (see Arnol'd 1963, Kolmogorov 1954, Moser 1973 and Siegel and Moser 1971). Unfortunately, the Solar System lies outside those ranges.

Large scale numerical integrations of the equations of motion by high speed computers are more promising but have their limitations imposed by error accumulation, storage and processing of data and imprecise starting conditions. The Cohen-Hubbard-Oesterwinter computation of the five outer planets over 10^6 years employing Cowell's method and a fixed step size of 40 days shed new light on the interactions of the outer planets, especially on the Neptune-Pluto locking mechanism. A 5×10^6 year numerical integration by Kinoshita and Nakai (1985) has recently been used by Milani and Nobili (1985) to display a further locking mechanism between the apses of Jupiter and Uranus. Two recent integrations of the orbits of the five outer planets, the ORRERY study (Applegate et al 1985) and Project LONGSTOP (report in preparation) for 2×10^8 and 10^8 years respectively have been completed and show that even for such long periods of time, no change in the hierarchy of the five outer planets' orbits takes place. The ability to perform such massive integrations signifies the rapid progress in recent years in the use of fast, large capaeity computer and data-handling techniques for problems of this nature. It is to be noted that 10^8 years is some 2% of the putative age of the Solar System. Such programmes of research will aid greatly our understanding of the dynamics of n-body HDS's, particularly the important question of the stability of the solutions to the equations of motion of such gravitationally-interacting systems. Important related problems to such studies are of course the question of the origin of planetary systems, their life-times and the frequency of occurrence in the universe of places suitable for life.

The c^2H criterion

It has already been remarked that only in the general 3-body HDS problem can the question of hierarchical stability be answered analytically in an absolute manner. In such a problem two of the masses form a binary system with the third in a larger, non-crossing orbit about the

centre of mass of the binary. If at any time the product of the total energy H and the square of the angular momentum c of the system is less than a certain value c^2H_{crit}, the system is hierarchically stable for all time; the third body will never disrupt the binary though it is possible that the third body will escape to infinity. The quantity c^2H_{crit} can always be computed from the three mass-values to obtain the criterion. Much work has been done on this problem in recent years (see for example Bozis 1976, Easton 1971 and 1975, Golubev 1968, Marchal 1971, Marchal and Saari 1975, Roy 1979, Roy et al 1984, Smale 1970, Szebehely and Zare 1977, Valsecchi et al 1980, Zare 1976, 1977).

For example, if we consider the system Sun-Jupiter-Saturn and calculate the c^2H criterion for it, we find that the hierarchical stability of that system is assured. All other three-body systcms with the Sun as one of the bodies and involving two planets are likewise stable, as are those systems consisting of a planet and two of its major satellites. An interesting set of systems comprising planet-satellite-Sun with the Sun disturbing the satellite in its orbit about the planet gives different results depending upon whether the Sun is assumed to move in a circular orbit about the planet-satellite centre of mass or is given its real orbital eccentricity (Valsecchi et al 1984). In this former case, all satellites except the outermost retrograde satellites of Jupiter and Saturn satisfy the c^2H stability criterion though the Earth's moon is a marginal case: in the latter case (solar orbital eccentricity non-zero) all but one (Triton) fail to satisfy the c^2H criterion for hierarchical stability. This surprising result probably shows how severe the criterion is rather than being an indication of instability of planetary satellites against solar perturbations.

Nevertheless, in the Solar System, no three-body subsystem can exist totally unperturbed by other bodies and the absolute stability criterion is therefore strictly speaking invalid. It is possible, however, to treat four-or-more-body systems found in nature as sets of disturbed three-body systems and two related approaches have been made from such a point of view.

The empirical stability parameter method

For $N \geqslant 4$, no analytical criterion valid for all time exists. The Roy-Walker empirical stability parameter approach attempts to identify parameters of the system that determine the probable subsequent hierarchical history of the system. The keyword is '*probable*'. A close analogy to the proeedure is the case of a planetary atmosphere where it is possibile

to calculate the half life T of the atmosphere as a function of a parameter x. In a time T, half the molecules in the atmosphere will have escaped from the planet. The quantity x is the ratio of the mean velocity of the atmosphere molecules to the velocity of escape. As x decreases, the half-life T, from being very short, gradually increases until, increasing ever more rapidly, it quickly reaches an astronomical duration. We note that while the half-life T for a given x can be exactly evaluated. The time at which any particular molecule will escape is uncertain. Nonetheless, for the atmospheres A and B with half-lives T_A and T_B, $T_A \gg T_B$, we can certainly be confident that it is overwhelmingly probable that a given molecule in A will remain far longer than any given molecule in B.

Roy and Walker's work on three and four body HDS's has identified empirical stability parameters, the so-called ϵ's, functions of the masses and ratios of their separation, the α's, which act as x does in the planetary atmosphere problem. All HDS's whose ϵ's and α's fall into a narrow range will have the same half-life T. The interpretation of T is that half of these systems will have changed their hierarchy (by cross-over of orbits, close encounters, escapes) after a time T. In a large number of numerical experiments, Roy and Walker have been able to draw half-life contours for three and four body systems which enable reliable predictions to be made for the duration of given HDS's. Although the work is incomplete, it is already possible to state that the planetary system and satellite systems in the Solar System must fall into ϵ and α ranges where the half-life is of astronomical duration (see Roy, 1979, Roy 1982, Roy et al. 1985, Walker et al. 1980, Walker and Roy 1983a, b) suggesting that it is highly probable that the Solar System not only has maintained its present hierarchical distribution of orbits for an astronomical time but will continue to do so for a future time duration equally long.

The Milani-Nobili method

Milani and Nobili (1983a, b) sought for a general perturbation theory relating the hierarchical stability lifetime of a four-body HDS to the rate of change of the absolute stability criteria c^2H of each of its three-body subsets as they are disturbed by the fourth body. Since the critical value of the function c^2H, viz. c^2H_{crit}, is a function only of the three masses, it is constant. While $c^2H < c^2H_{\text{crit}}$, for a given three-body subset, the subset remains hierarchically stable. Milani and Nobili's approach, which in its analytical development makes use of the Roy-Walker empi-

rical stability parameters, provides a means of calculating the minimum time perturbations will take to increase c^2H to c^2H_{crit} for a given three-body subset. For example, in the four-body system Sun-Mcrcury-Venus-Jupiter, they conclude that the hierarchy of the subset Sun-Mereury-Venus is stable, against perturbations by Jupiter, for at least 1.1×10^8 years while the subset Sun-Venus-Jupiter, against perturbations by Mcrcury, is stable for at least 3×10^9 years.

Near mirror configurations as a stability mechanism in the solar sistem

In 1955 Ovenden and Roy published the mirror theorem which states that if n point masses obeying Newton's law of gravitation achieve at a particular epoch a configuration where all the mutual velocity vectors are perpendicular to all the mutual radius vectors, then their trajectories after that epoch are mirror images of their trajcctories before that epoch. Only two distinct mirror configurations can exist, one where all the bodies are collinear with their velocities perpendicular to that line, the other where all the bodies occupy a plane with their velocities perpendicular to that plane.

The corollary to the theorem is obvious. If mirror configurations occur at two epochs then the system is periodic. This says nothing about the stability of the system. Left to itself, the system will last forever even though close encounters of bodies in the system occur. But, if additional perturbations occur (for example an intruding body passes by the system), even tiny perturbations occurring at a time when a close encounter is taking place will probably disturb the system irrevocably leading to escape, collision or change of hierarchy. Ovenden and Roy argued that to ensure stability, the time intervals between mirror configurations should be as small as possibile since during a time interval, disturbances build up in semimajor axes, eccentricities and other elements which will only be cancelled during the next time interval. If such disturbances are allowed to propagate for too long a time, disaster may intervene before reversal takes place.

In a recent paper Roy (1983) re-examined the question of the conditions conducive to the occurrence of close and frequent approximations to mirror configurations in the Solar System. For hierarchical stability the basic need is that mutual perturbations in semimajor axes and eccentricities should be reversed before they grow big enough to threaten the stability of the system. The optimum conditions are that the members of the system move in well-spaeed out orbits of small eccentricity

and inclination, with small Roy-Walker empirical stability parameters. The argument is simple and may be illustrated by application to the planets Jupiter and Saturn. Even for them their stability parameters are so small ($< 2.5 \times 10^{-4}$) that the synodic period of their apses (the time between successive similar geometrical configurations of the apses'heliocentric longitudes) is very long (of order 50,000 years) compared to the planets' synodic period (of order 20 years). Inevitably therefore, whenever the apses are near conjunction or opposition every 25,000 years, there occur many conjunctions and oppositions of Jupiter and Saturn. For example every time the apses are within 1° of each other, almost 50 conjunctions and oppositions take place ensuring that one within a few degrees of the apses will occur. The small eccentricities and mutual inclination of the orbits, together with the small true anomalies at such a time produce an almost perfect mirror configuration (to within 12 arc min in the example quoted). Thus at every possible opportunity an almost perfect reversal of Jupiter and Saturn's perturbations on each other occurs. It is obvious that if the disturbing forces were higher (i.e. larger stability parameters) the apses would rotate faster, the eccentricities would be larger so that during the shorter intervals of time when the apses were near each other, fewer conjunctions and oppositions of the bodies would occur; moreover, in order to get as 'good' a near mirror configuration with higher eccentricities, the conjunction or position would have to be at smaller true anomalies thus radically lowering the probability of success. The scenario for disaster is a system with high perturbations so increasing eccentricities and changing semimajor axes that when the apses come together the probability of bodies succeeding in coming into a conjunction or opposition of small enough true anomaly to be an effective mirror is less than the probability that before that happens the perturbations produce a close encounter of bodies or a cross-over of orbits.

Conclusions

In nature, not only in the Solar System but in multiple star systems such as Castor, n-body systems are found in hierarchical dynamical structures. This in itself indicates that such structures are more stable than others. In the Solar System, both in the planetary system and the major satellite systems, the major bodies, well-spaced out, almost circular, almost coplanar orbits afford the greatest chance of long-term hierarchical stability according to the researches described in the earlier sections.

Although in the absence of analytical stability relations for HDS's of four and more bodies there can be no absolute pronouncement on the long-term stability of such systems, it is becoming clear that probabilistic life-times may be computed for them. Indications so far place the Solar System hierarchical dynamical systems among hierarchical stability lifetimes of an astronomical duration.

REFERENCES

[1] Applegate, J.H., Douglas, M.R., Gursel, Y., Sussman, G.J. and Wisdom, J.: 1986, The Astron. J., **92**, 176.

[2] Arnol'd, V.I.: 1963, Usp. Mat. Nauk (Sov. Math-Usp) **18**, No. 6 91: No. 5, 13.

[3] Bozis, G.: 1976, Astrophys. Sp. Sci. **43**, 355.

[4] Cohen, C.J. and Hubbard, E.C.: 1964, Naval Weapons Laboratory Report No. 1945.

[5] Cohen, C.J.and Hubbard, E.C.: 1965, Astron. J., **70**, 10.

[6] Cohen, C.J., Hubbard, E.C. and Oesterwinter, C.: 1967, Astron. **72**, 973.

[7] Cohen, C.J., Hubbard, E.C. and Oesterwinter, C.: 1972, Astron. Pap. Am. Ephem. **22**.

[8] Easton, R.: 1971, J. Diff. Eq., **10**, 1971.

[9] Easton, R.: 1975, J. Diff. Eq., **19**, 258.

[10] Golubev, V.G.: 1968, Doklady Akad. Nauk. SSSR **180**, 308.

[11] Kinoshita, H. and Nakai, H.: 1985, Celest. Mech. (in press.).

[12] Kolmogorov, A.N.: 1954, Doklady Akad. Nauk. SSSR (Sov. Phys. — Doklady) **98**, No. 4.

[13] Marchal, C.: 1971, Astron. Astrophys., **10**, 278.

[14] Marchal, C. and Saari, D.: 1975, Celest. Mech. **12**, 115.

[15] Milani, A. and Nobili, A.: 1983a, Celest. Mech. **31**, 213.

[16] Milani, A. and Nobili, A.: 1983b, Celest. Mech. **31**, 241.

[17] Milani, A. and Nobili, A.: 1985, Celest. Mech. **35**, 279.

[18] MOSER, J.: 1973, *Stable and Random Motions in Dynamical Systems*, Princeton University Press, Princeton, N.J.

[19] ROY, A.E.: 1979, «Empirical Stability Criteria in the Many-Body Problem», in *Instabilities in Dynamical Systems*, ed. V. Szebehely, Reidel.

[20] ROY, A.E.: 1982, «The Stability of N-body Hierarchical Dynamical Mechanics», in *Applications of Modern Dynamics to Celestial Mechanics and Astrodynamics*, ed. V. Szebehely, Reidel.

[21] ROY, A.E.: 1983, «Asymptotic Approach to Mirror Conditions as a Trapping Mechanism in N-body Hierarchical Dynamical Systems» in *Dynamical Trapping and Evolution in the Solar System* ed. V.V. Markellos and Y. Kozai, Reidel.

[22] ROY, A.E., CARUSI, A., VALSECCHI G. and WALKER, I.W.: 1984a, Astron. Astrophys. **141**, 25.

[23] ROY, A.E. and OVENDEN, M.W.: 1955, Mon. Not. Roy. Astron. Soc. **115**, 296.

[24] ROY, A.E., WALKER, I.W. and MCDONALD, A.J.C.: 1985, «Studies in the Stability of Hierrchical Dynamical Systems» in *Stability of the Solar System and its Minor Natural and Artificial Bodies*. ed. V. Szebehely, Reidel.

[25] SIEGEL, C.L. and MOSER, J.K.: 1971, *Lectures on Celestial Mechanics*, Springer-Verlag, Berlin, W.-Germany.

[26] SMALE, S.: 1970, Invent. Math. **11**, 45.

[27] SZEBEHELY, V.G. and ZARE, K.: 1977, Astron. Astrophys. **58**, 145.

[28] VALSECCHI, G.B., CARUSI, A. and ROY, A.E.: 1986, Celest. Mech., **32**, 217.

[29] WALKER, I.W.: 1983, Celest. Mech. **29**, 149.

[30] WALKER, I.W., EMSLIE, A.G. and ROY, A.E.: 1980, Celest. Mech. **22**, 371.

[31] WALKER, I.W. and ROY, A.E.: 1983a, Celest. mech., **29**, 117.

[32] WALKER, I.W. and ROY, A.E.: 1983b, Celest. Mech., **29**, 267.

[33] ZARE, K.: 1976, Celest, Mech. **14**, 73.

[34] ZARE, K.: 1977, Celest. Mech., **16**, 35.

Orbital Stability and origin of the Minor Bodies of the solar system

Raimundo O. VICENTE (*)

1. Introduction

This international symposium has been organized in order to pay tribute to the late Prof. G. Colombo. He was of the scientists who applied the known concepts of celestial mechanics to develop original ideas to the motions not only of the components of the solar system but also to space probes. We are explaining some ideas about the stability of the minor bodies of the solar system and infer, from their stability, some consequences about possible evolutionary lines for their behaviour.

One of the features of scientific papers about the minor bodies of the solar system is the presentation of statistics giving a wealth of information about their different features. This procedure is partly justified by the fact that we are dealing with the most numerous components of the solar system and, therefore, it might be possible to use statistical distributions in order to infer certain properties common to all these bodies.

We are interested to see if it is possible to infer some evolutionary behaviour of these components,due to the fact they are so numerous. One cannot apply the same type of statistics to infer the evolutionary stability of the planets and their satellites because they are so few.

We are specially interested in the evolutionary trends of the minor bodies during the main stage of the total duration of the solar system. For this purpose it is convenient to divide the whole time span of the existence of the solar system in 3 stages: 1) initial; 2) main; 3) final.

It is well known that the durations of the initial and final stages depend on the cosmological theories adopted, but I think it is a fair assumption to say that the main stage will have a far longer time span than the other 2 stages. We are admitting values of the order of 4×10^9 years.

(*) Dept. Applied Mathematics, Faculty of Sciences, Lisbon, Portugal.

2. Statistical facts

It is possible to gather some values about the number and orbital features of the minor bodies.

Table 1

	Approximate total number	Orbital elements a (A.U.)	i°	Known orbits
A. ASTEROIDS	30,000			~ 3,500
$d \geq 130$ km	100			
Trojan	Several hundreds			
A1. Trans-Jovian		> 5	6-42	20
A2. Main asteroidal belt Nine families		2.16-3.07	0.5-27	860
A3. Crossing Earth's orbit	1300		≤ 64	100
Aten	100	< 1		5
C. COMETS	10^{12}			600
C1. Short period			3-162	~ 100
C2. Long period (P > 200yrs)			≤ 178	500
M. METEOR STREAMS	100	1—59	< 173	~ 50

An inspection of the data presented leads us to some useful conclusion for our purposes. First of all we see that the number of known orbits constitutes a very small sample in comparison with the forecasts of the total number of asteroids and comets and, therefore, we have to be cautious about infering conclusions from such a small sample. In the case of meteor streams in the vicinity of the Earth, we have really few in comparison with the numbers of asteroids and comets, and that might be linked to their origin and evolution during the main stage of the duration of the solar system.

Looking at the inclinations of the orbits, we see they show a wide range of values for all the minor bodies in comparison with the planets, and the same conclusion applies for the eccentricities. The semimajor

axis also shows a wide range from comets very far away from the planets down to values well inside the Earth's orbit for some asteroids and meteor streams.

This simple Table 1 shows one of the main features of the minor bodies, that is, the researches referring to their motions should be done in 3 dimensions, whenever possible, in order to arrive at more valid conclusions in agreement with the spatial distribution of these bodies. The asteroids in the main belt (A2), in spite that they do not exhibit such a wide variation of their orbital elements when compared with the other 2 groups (A1 and A3), nevertheless they show such a range of inclinations that the problem of explaining the Kirkwood gaps should be treated in 3 dimensions.

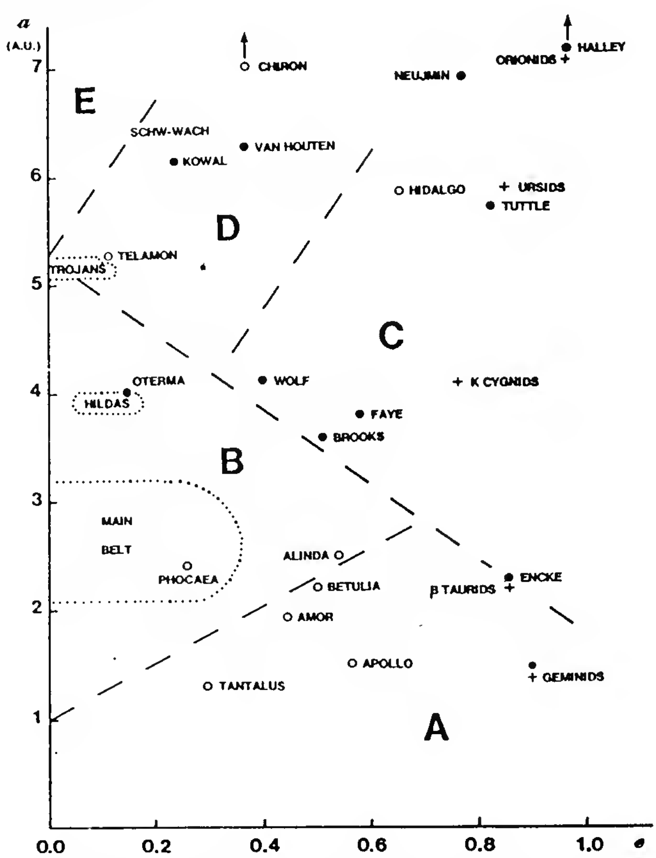


Fig. 1 - Some asteroids (open circles) short period comets (solid circles) and meteor streams (crosses) plotted in a diagram of semi-major axis versus eccentricity.

We are considering a diagram (Fig. 1) in two dimensions of the semi-major axis versus eccentricity that might give us some hints about the origin and relationships among these minor components of the solar system. We are doing this in spite of our warning that the orbital problems of the minor bodies should be treated in 3 dimensions.

A quick look at this diagram shows us that the distribution of the minor bodies is not a random distribution and there are fairly well defined areas occupied by different types of bodies: in regions A and B appear a predominance of asteroids while the influence of Jupiter appears on the orbits of regions C and D, being noticeable the existence of comets and some of their associated meteor streams in region C; region E corresponds to the area of the trans-jovian minor bodies.

If we had introduced the inclination, we would notice that the minor bodies in regions C, D and E have higher inclinations than in region B and that might be linked with the stability of their orbits from the point of view of evolution.

Looking at regions A and C, we notice that meteor streams are situated in these regions and, also, there seems to exist an association between the location of meteor streams and some comets.

There are many diagrams in two dimensions trying to prove different theories and hypotheses, depending on the authors who have plotted them. But the main point, that should not be overlooked, is the fact that the minor bodies have definitely a 3 dimensions distribution contrary to what happens with the planets.

3. Orbital stability

The problem of the stability of the orbits of the various components of the solar system has been one of the fundamental questions of celestial mechanics since its beginnings. Laplace was one of the first scientists to be interested in the stability of planetary orbits because he was interested in the origin and evolution of the solar system.

This fundamental problem raises several questions, not only about the adequate definitions of stability but also the time intervals we are considering, in order to see that the stability is maintained throughout the whole system or applying only to certain components. We shall examine a few aspects of this delicate and important question.

We are interested in the orbital stability of the minor bodies referring to the main stage of the evolution of the solar system, that is, we restrict ourselves to the far longer period of the existence of the planetary system,

and our considerations are not concerned with the initial and final stages of the system. It is possible to forecast that the behaviour of the minor bodies is different from the planets during this stage because of the greater variety of their orbital elements.

Another advantage of the researches about the evolution of the minor bodies is the fact that they are so numerous and, therefore, will offer us greater chances of detecting variations in their orbital stability.

Let us now consider another important aspect of the question of stability. It refers to the precise definitions of stability and we should like to formulate some questions on the subject:

What do we mean by stability? Is it independent of the time scale? Are the known definitions of stability equivalent? Do they refer to the same concepts? Is the mathematical formulation of stability adequate only for a few centuries? Are our mathematical models good enough to represent the stability during the main stage of the evolution of the system?

I think it is easier to formulate these questions than to find the adequate answers.

We are interested in the definitions of stability which have a practical and immediate application to the case of the minor bodies of the solar system. This narrows the field of suitable definitions, and we can consider the following ones:

1) The measure of stability *S* introduced by Szebehely (1978) and based on Hill's concept of stability. This definition has two great advantages: a) it is an analytical definition and, therefore, it does not depend on numerical techniques and its pitfalls about convergence and rounding off errors; b) it can be applied to 3 dimensions and, therefore, very suitable for the minor bodies with their great variety of inclinations. We can determine the measure of stability *S* for any body of the solar system thanks to a number of tables that have been computed by Szebehely et al. (1983).
2) Some interesting stability results can be derived by the method of surface of section, revealing the main features of the phase space; this technique has been refined by Hadjidemetriou and Ichtiaroglou (1984) since they employed analytic results for the determination of the energy levels at wichi the surfaces of section are taken. This is a different approach and, therefore, it is convenient to compare with the previous definition of stability.
3) The determination of Liapunov's Characteristic Numbers (LCN) derives from the definition of stability established by Liapunov and which

is connected with the concept of isochronous correspondence. This approach to the definition of stability is based on numerical integration and, therefore, quite different from the previous criteria. In a way it is useful that this criterion has a different theoretical approach so we can compare the results obtained from several view points. Being a numerical method, we have to be careful about the procedures of numerical integration, and it has the disadvantage of being necessary many integration steps before we can discriminate between stability and instability.

4. Asteroids and Kirkwood gaps

Let us follow the historical developments of the discoveries of these bodies. The first ones to be observed were evidently the bigger ones, with greater masses, and they were called minor planets because their orbits have, as primary, the Sun. The number of so-called minor planets, with diameters bigger than 130 km, is about 100, that is, a very small number in comparison with the total number of these bodies which statistical estimates evaluate as about 30,000 accessible to our present day instruments of observation (Table 1).

We have here already a selection effect because the great majority of these bodies have small dimensions, like, for instance, 2km in diameter for one of th Aten group. But we can still consider the broad band of debris which probably resulted from collisions, having still smaller masses, and orbits around the Sun. Can we call them minor planets?

Considering the overall features of these small bodies of the solar system, I think the name asteroids is more appropriate. As you see, this subject is so interesting that there is not a consensus among scientists about the correct name to designate them.

Besides the wide range of masses of the asteroids, we also have a great variation of the inclinations and eccentricities of their orbits. Let us try to deduce some relationships between the morphological aspects of the orbits and the origin and evolution of the asteroids. There are many ways of classifying them, and we adopt the following classification:

A1 - Asteroids beyond Jupiter: there is a wide range of inclination, but there are not many known so far because of observational difficulties

A2 - Main asteroidal belt: the elements of their orbits are more similar among them than A1 or A3 but even so they show a three dimensional distribution

A3 - Earth's crossing orbits: there is a variety of inclinations but they are very difficult to observe and, therefore, we have an important selection effect. Among them we should mention the Aten group with only about 6 observed and a statistical approximate number of about 100.

This simple grouping of the asteroids in 3 classes raises already several questions: Why is class A2 far more numerous than A1 and A3? Why A1 and A3 present greater variations in their inclinations? Is it valid to say that class A2 has a different evolution during the main stage than classes 1 and 3?

We can also postulate that all the asteroids had the same origin but, due to unknown causes, some of them evolved in such orbits that originated classes A1 and A3. But we can also think that asteroids A1 and A3 did not originate and evolved from class A2 asteroids.

Let us now apply the criteria of stability we have mentioned before and see the possible behaviour of their orbits.

The calculation of the measure of stability S gives the following results: we obtain $S < 0$ for classes A1 and A3, that is, they show instability in agreement with Hill's concept while we compute $S > 0$ for the majority of asteroids of the main asteroidal belt (class A2).

The computation of Liapunov's Characteristic Number (LCN) for some of the asteroids in classes A1, A2 and A3 gives us results in agreement with the previous ones. The determination of the convergence towards zero, meaning that the asteroid presents stability, is a numerical procedure that needs thousands of steps in the integration and, therefore, cannot be applied easily to so many asteroids.

The application of the method of surface of section, employing the model of the planar circular restricted 3 body problem with the Sun and Jupiter as primaries, to the stability of the asteroids also gives results in agreement with the previous criteria.

This is an interesting conclusion about the stability of the asteroids, which can be grouped in 3 classes such that classes A1 and A3 show in stability while class A2 shows greater stability. We should have some confidence in these conclusions because we obtained these results from 3 different criteria of stability.

Our stability calculations are in agreement with the conclusions we had derived from the morphological aspects of the orbits, and we can say that orbits with higher inclinations and greater eccentricities lead to greater instability. In this way the existence of fewer asteroids in classes A1 and A3 than in class A2 is justified by the fact that their orbits tend to evolve faster than the ones in the main asteroidal belt.

We can therefore say that there are two distinct types of evolution during the mains stage of the solar system: one corresponding to the main asteroidal belt (class A2) where the evolution times will be similar to the planets, another corresponding to classes A1 and A3 showing a quicker evolution and so there are fewer members of these asteroidal groups. We should classify the first part of this statement in the sense that the variation of masses of the main asteroidal belt is so great that we should put a lower limit on the dimensions of the asteroids that will behave like the planets; considering smaller masses, for instance, what is called the asteroidal debris, their evolutionary behaviour will be quite different and they will certainly have orbits that will evolve in a quite different time scale. But the question is to know what is the limiting value of the dimensions of the asteroid that will make the transition from one type to the other type of evolution. Is it a diameter of 1 km or 0,1 km?

We have not yet enough detailed observations to answer this question and we are always hampered by selection effects due to the fact that the great majority of observed asteroids are the easier ones to be detected.

The problem of the origin of classes A1 and A3 is more difficult than in the case of the main asteroidal belt because we can consider at least 3 hypotheses: 1) their orbits evolved from orbits in the main asteroidal belt; 2) their orbits are contemporary to the orbits of the main asteroidal belt; 3) their origin is quite different and it might be associated with some comets.

All these hypotheses raise interesting questions and it has not yet been possible to find a satisfactory solution. For instance, in the first hypothesis it is required some very particular initial dynamical conditions for the orbits to change so drastically the values of a, e and i. In the second hypothesis we have to consider some peculiar conditions in the initial stage of the evolution of the solar system that put those orbits with such values of their elements, specially high inclinations and great eccentricities. In the third hypothesis we have to find out an adequate relationship between these asteroids and some special group of comets.

We have seen that classes A1 and A3 have far fewer members than class A2, and in order to explain this fact we can consider at least 3 possibilities: a) they are very few because their orbits are unstable and therefore they will have a faster evolution during the main stage; in this case we have to consider that they will have a limited duration and this type of asteroids will disappear from the solar system; b) still considering that their orbits are unstable we have to discover some sort of process that will replenish these classes A1 and A3 with new asteroids in order that we can consider them as permanent members of the solar sy-

stem during the main stage of evolution; c) another possibility would be to consider their orbits as stable, after they evolved in such high inclinations and great eccentricities, and therefore, they will last forever during the main stage. But this possibility is not in agreement with our statements about the instability of their orbits derived from different stability criteria.

Looking at the literature on the subject of evolution of the solar system, we can notice that a vast majority of papers consider all the components of the solar system (planets, satellites and minor bodies) as permanent members, that is, since their formation, during the initial stage, they will last forever until the final stages of the system. It is rare to find ideas and hypotheses mentioning that, at least for some of the minor bodies, their evolutions are faster and, therefore, they will disappear during the main stage of the evolution. This hypothesis, a priori, is as good as the other and we are not breaking any physical laws saying that some components of the solar system, for instance, some classes of asteroids (Earth's crossing) have a limited duration in the planetary system and they might disappear forever.

We can now consider another interesting subject connected with asteroids of class A2, that is, the main asteroidal belt. It refers to the existence of gaps and they are generally called Kirkwood gaps.

The Kirkwood gaps might be an interesting example of selection effects which bedevils the subject of the minor bodies. As we have seen, in a real perspective of the problem, we must consider 3 dimensions, but, besides that, we have to define the masses and brightnesses of the asteroids we can consider valid for defining the dimensions of the gaps.

If we had enough accurate observations of the asteroids, we might arrive at the conclusion that the Kirkwood gaps are not so well defined as it appears at the present time with our small sample of data. As we know, considering asteroids of greater magnitudes, that is, with smaller dimensions, the main asteroidal belt becomes more crowded and it will appear that the gaps will not be so sharpely defined. We might, therefore, infer that, besides the main resonance effects, there are other causes, acting specially on the asteroids of smaller dimensions, that will put them in the so-called forbidden regions of the gaps.

The subject has been studied in hundreds for papers, usually in two dimensions. Due to the fact that asteroids show greater variations in the values of their elements permits the setting up of more elaborate theories, for instance, the analytical modelling of high eccentricity librations and its application to certain groups of asteroids (Hildas and Thule), existing in orbital resonance with Jupiter, and situated in the wide re-

gion between the outer boundary of the main asteroidal belt and the orbit of Jupiter (Ferraz-Mello, 1988).

We can compare the asteroidal belt with their gaps to the similar morphological aspect of the rings and their associated satellites in the case of the bigger outer planets, namely Saturn. We have there a complex distribution of particles forming the rings with some gaps originated by resonances due to satellites. We can contemplate a similar morphological case for the asteroids in the main belt, that might be considered as a ring of particles around the Sun, and with their associated gaps provoked by resonances of the planets Jupiter and Mars. We can even speculate that some of the narrower rings have «shepherd» asteroids, namely the bigger ones, skirting inside and outside each ring, and their subtle gravitational pull is supposed to force the ring material to assume narrow shapes.

If we consider the total dimensions of the solar system in relation to the dimensions of the system of rings, called the asteroidal belt, we have a similar morphological aspect to the system of planetary rings, which could be well seen by an observer placed far away from the Sun. At such distance, the thickness of the asteroidal ring would appear negligible, but this is not the case for an observer situated on the Earth.

We can nowadays begin to distinguish between the asteroids and their debris, forming probably thin rings, which would permeate throughout the entire asteroidal belt. Following the morphological analogy, we could say that the solar system of rings is the so-called main asteroidal belt.

We have therefore 2 pictures, one for an observer far away from the Sun and another for a terrestrial observer and, in this case, we have to consider the rings and gaps in 3 dimensions and, also, the asteroids of very small dimensions which we call asteroidal debris. Here, again, we have a problem of definition because all of them have orbits around the Sun, but their dimensions are so small that, from the dynamical point of view, they should probably be given a different name (debris) and analytical treatment.

5. Comets

We are confronted with the same problem of a very small sample of known comets in comparison with the statistical guesses about their total numbers. Also the great majority of known orbits correspond to the long-period comets with periods greater than 200 years. The better studied ones are the short period but we only know about 100 and that is a very small number.

We have noticed that they appear in regions C and D of the diagram (Fig. 1) and their orbits show greater inclinations and eccentricities than other minor bodies of the solar system. The computation of the measure of stability S leads to negative values meaning they present greater instabilities in their orbits.

We know that comets loose mass whenever they approach the Sun and, therefore, it is a valid scientific inference to say that comets are short-lived components of our system, specially the short period ones. I think that this statement is one of the few statements about comets that have the agreement of the researchers on the subject.

In order to explain the existence of short and long-period comets we can consider two hypotheses: 1) both have the same origin; 2) the origin of the short period comets is different from the long period comets. Both these hypotheses can be upheld and the main reason for that is the lack of sufficient observations, so we have not enough data to discriminate between them.

Once we admit that the life times of comets are limited during the main stage of the evolution of the system, we have to discuss the behaviour of the end product of this evolution. We can consider two cases: 1) they disappear because they loose their masses completely; 2) the end products of comets might originate: a) certain types of asteroids (classes A1 and A3); b) meteor streams.

Let us consider the second case in greater detail because it will have more consequences for the evolution of some minor bodies during the main stage of the solar system. If they evolve into asteroids we have to think about the dynamical initial conditions necessary for producing asteroidal orbits beyond Jupiter (class A1) and crossing the Earth's orbit (class A3); these initial conditions are more likely to be satisfied for class A3 because that corresponds to the region of perihelion passages of comets and, therefore, of greater perturbation of the physical and dynamical parameters of the comets, originating asteroids of the Apollo-Amor groups and Aten group. But, of course, we can also imagine that the perturbations are such that the orbits of this end-products of comets become trans-jovian.

We can see that comets are important components of the planetary system, independently of the hypotheses adopted for their evolution, because they originate other types of minor bodies and they can be considered as sources of distribution of mass in different regions of the solar system. This is quite unique for any component of the system due to the fact of their orbital instability and physical structure. For instance, we can observe comet dust trails, formed by particles up to the centimeter-

size, and they are spread along the orbits; it is even postulated that some dust trails might correspond to the existence of comets not yet discovered.

It has been verified from statistical studies of cometary orbits that they seem to originate from a very localized region of the celestial sphere. The original researches were made by Oort (1950), and I should like to call this fact as Oort's effect, without any implication about the origin of comets because that is a different problem.

Once we agree that some cometary orbits are unstable, we have to examine possible hypotheses to explain their origin and evolution. We can consider, at least, two hypotheses:

1) they are permanent members of the solar system since its origin until the final stages.

In this case we have to set up a theory implying their formation from the initial stages of the solar system, and that there is enough mass to keep them forever, because we know that individual comets have a limited duration.

One theory considers the comets as constituting a population of planetesimals that have been ejected to the outer regions of the solar system because of perturbations in their orbits originated by the main planets. In order to explain the fact that they appear from a limited region of the celestial sphere (Oort's effect), it is considered that they form

a cloud, of enormous mass, which corresponds to a huge reservoir of comets that it will produce comets until the final stages of the solar system. This cloud is called Oort's cloud by some specialists.

2) they are not permanent members.

In this case we can consider a similar theory for their origin, admitting either the existence of a cloud or several clouds, depending on the evolution and duration we would like to think for the comets. If there is only one cloud, the comets will disappear forever, as members of the solar system, when the mass of that cloud is exhausted. If there are several clouds in the trajectory of the Sun around the centre of mass of the galaxy, then we shall see comets whenever we have suitable clouds. But, in any case, the comets will only exist until the mass of these different clouds is exhausted. In this theory we do not need to imagine a cloud with a huge mass, and very special initial dynamical conditions all the time, which are necessary for the permanent production of comets.

6. Meteor streams

There seems to be a consensus among researchers on the subject that these minor bodies are not very numerous and, at least some of them have a close connection with certain types of comets.

The computed values of the measure of stability S gives negative values not only for the meteor streams but also for the parent periodic comets.

We are very much hampered by selection effects in their observations because it depends on the geometrical configuration of the position of the terrestrial observatory in relation to the orbit of the meteor stream. We only know about 100 and we can see in Fig. 1 some of the short-period comets associated with meteor streams.

If we adopt the theory that all meteor streams are originated from comets, then we can state that they are not original components of the solar system, because the first meteor streams could only by formed after comets evolved their orbits around the Sun in such a way that they became unstable. We could see meteor streams only when this orbital instability appeared and, therefore, after the initial stage of the planetary system had entered the main stage.

Another statement we can make about them refers to the fact they are really the only minor bodies, where we have observed their orbital elements vary, during the short interval of accurate astronomical observations.

We can also say they are not very numerous because, after their separation from the parent comets, their orbits are so unstable that they have a very short life time in comparison with the main stage of the evolution of the solar system. Admitting this theory, we can state that there will always be a very limited number of meteor streams around the Earth.

If we consider the existence of meteor streams associated with some types of periodic comets, we can infer that there will be meteor streams associated with other planets, specially the terrestrial planets like Mars and Venus. But we can also consider probable the existence of meteor streams around Jupiter, and this statement might be supported from the number of cometary orbits that are influenced by Jupiter.

7. Conclusions

We are interested about the evolution of the orbits of the minor bodies during the main stage of the life of the solar system, and we arrived

at the following conclusions:

1) the meteor streams have the shortest duration during that time interval
2) the existence of short period comets, depends on the theories adopted for their origin
3) the orbits of Earth's crossing asteroids show instability and, therefore, they will have a tendency for shorter evolutionary lives
4) the asteroidal debris will present an orbital evolution different from the bigger and main asteroidal belt, because they are subject to physical processes affecting their motions, for instance, collisions
5) the bigger asteroids, in the main asteroidal belt, will present an orbital evolution in the time scale of the main stage.

The comets are very important for the evolution of other minor bodies and they can be considered as building blocks for the origin of asteroids in regions A (Earth's crossing orbits) and E (beyond Jupiter's orbit) and meteor streams. Their origin and evolution can be explained by two theories: one admitting they are permanent members of the solar system and the other considering they have a limited duration, and, therefore, they will not exist until the final stages.

Let us hope that our great lack of astronomical observations will be much improved in the next decade with the setting up in orbit of the Space Telescope, and the Galileo space probe whose mission includes a voyage among the asteroids for nearly a year. We can then verify or reject some of the theories and hypotheses put forward in the previous paragraphs.

REFERENCES

FERRAZ-MELLO, S., 1988 A.J., **96**, p. 400.

HADJIDEMETRIOU, J.D. and ICHTIAROGLOU, S., 1984 Astron. Astrophys. **131**, p. 20.

OORT, J.H., 1950 Bull. Astron. Inst. Neth., **11**, p. 91.

SZEBEHELY, V., 1978 Cel. Mechanics, **18**, p. 383.

SZEBEHELY, V., VICENTE, R.O. and LUNDBERG, J., 1983, «Dynamical Trapping and Evolution in the Solar System», V.V. Markellos and Y. Kozai (eds.), D. Reidel Publ. Co., p. 123.

Basi osservative della relatività generale

Tullio REGGE

Ci troviamo riuniti per commemorare Giuseppe Colombo, un grande Maestro ed una delle figure più luminose e rappresentative della Scienza italiana contemporanea. Colombo era un conoscitore profondo della meccanica celeste ed a lui si debbono molte brillanti intuizioni da cui ha tratto grande profitto l'esplorazione del sistema solare. In questo egli continuò l'opera di Newton di Laplace e di Lagrange e la sua ricerca ritiene un indirizzo ed una impronta classica in cui vennero portati fino alle estreme conseguenze i metodi della meccanica celeste.

Senza una profonda conoscenza di questa non sarebbe stato possibile tuttavia misurare e valutare appieno quelle minime deviazioni dalla legge di gravitazione universale che richiedono, per una corretta spiegazione, l'uso della relatività generale.

Questa seconda versione della teoria della relatività, versione che contiene in sé come caso particolare quella ristretta, fu costruita da Albert Einstein durante la prima guerra mondiale e pubblicata nel 1916. Essa conteneva concetti di portata rivoluzionaria e proprio per questa ragione ebbe bisogno di controlli osservativi stringenti per essere accettata appieno.

Per circa mezzo secolo e cioè fino all'inizio dell'era spaziale essa si appoggiò su tre test classici che esaminerò brevemente.

1. La precessione del perielio di Mercurio, pari a circa 43'' d'arco per secolo su di un totale di oltre 5000'' dovuti alla perturbazione dei pianeti, era già nota a Leverrier il quale tentò di spiegarla attribuendola a Vulcano, un ipotetico pianeta vicinissimo al Sole ma che non fu mai visto. La teoria della relatività generale rimane in accordo quasi perfetto con questo dato. Usando una strumentazione più moderna ed anche misure dirette mediante radar gli astronomi hanno determinato i valori della precessione anche per gli altri pianeti e per l'asteroide Eros trovando una serie di risultati in ottimo accordo con la teoria. Nell'effettuare questi confronti è stata necessaria una revisione totale dei dati in nostro possesso riguardanti l'intero sistema solare. Questi successi non hanno eliminato le polemiche. Se il Sole fosse lievemente schiacciato ai poli e possedesse quindi un momento di quadrupolo ne risulterebbero alcune

perturbazioni sul moto di Mercurio per cui l'accordo con la teoria non sarebbe così perfetto. In questo caso ci sarebbe spazio per una generalizzazione della relatività detta di Iordan-Brans-Dicke in cui appare un campo scalare extra ma che non gode del favore della maggioranza dei fisici teorici e che comunque è ormai esclusa da altri dati.

2. La deflessione dei raggi luminosi sul bordo del Sole fu misurata per la prima volta nel 1919 in occasione di una eclisse di Sole da una spedizione inglese all'isola di Principe a cui partecipò Sir Arthur Eddington. Essa trovò dati in accordo con la teoria, con un errore di circa il 10%. Le misure furono ripetute durante altre eclissi di Sole ed infine osservando mediante radiotelescopi l'occultazione della radiosorgente 3C273 da parte del disco solare. In quest'ultimo caso il margine di errore è sceso a circa 1%.

Con altre tecniche si dovrebbe scendere sotto questo limite. Al tempo di Eddington la notizia destò un profondo stupore e senso di ammirazione per Einstein che diventò famosissimo. Essa rappresenta effettivamente una chiara deviazione dalla meccanica newtoniana che prevedeva un valore pari a metà di quello osservato.

3. Lo spostamento verso il rosso delle linee spettrali emesse da una stella, in particolare dal Sole. L'effetto esiste ma è mascherato dall'altro effetto Doppler generato dalla agitazione termica alla superficie della stella per cui occorre una certa cautela ed immaginazione per vederlo. Esso è stato comunque osservato sia sul Sole che su altre stelle, in particolare sulle stelle nane bianche in cui esso è particolarmente evidente. Infine esso è stato misurato direttamente sulla Terra e su raggi gamma da Rebka e Pound usando l'effetto Mossbauer. Come ultima conferma si sa che questo effetto è direttamente collegato e praticamente equivalente al rallentamento degli orologi in un campo gravitazionale che è stato pure rilevato, tra gli altri dal nostro Istituto Galileo Ferraris.

I tre effetti sopra descritti hanno costituito per anni il solo punto di appoggio della teoria. Ad essi si sono gradualmente aggiunti altri indizi ed altre conferme, tra di queste la misura dei ritardi nei segnali radar riflessi dai pianeti o da sonde lanciate nel sistema solare ed alcune anomalie presenti nel moto della Luna.

Tutte queste misure presuppongono una conoscenza sempre più dettagliata del moto dei corpi entro il sistema solare, questa dipende dalla conoscenza dei dati geometrici e dalle masse dei costituenti. Questo tipo di indagine è stato reso possibile, oltre che dalla esplorazione dello spazio, dall'uso massiccio di grandi calcolatori. Altre conferme potranno venire dal lancio di satelliti circumterrestri in cui si tenta di evidenziare sottili effetti di interazione tra il momento angolare del satellite ed il campo gravitazionale terrestre.

Molto più interessanti a questo riguardo appaiono le prospettive extrasolari. Negli ultimi decenni sono stati scoperti alcuni sistemi doppi composti da oggetti compatti (nane bianche, stelle a neutroni o pulsar, buchi neri) in orbita molto stretta. Uno di questi, PS 1916 + 13, consiste di due componenti con periodo orbitale di circa 8 ore e la cui orbita potrebbe essere contenuta entro il Sole.

Si hanno buone ragioni per credere che ambedue le componenti siano stelle a neutroni e che l'unica forma di interazione significativa sia quella gravitazionale. Il pulsar binario constituisce quindi un laboratorio ideale.

Gli astrofisici ben conoscono una ambiguità che sorge tutte le volte che si osservi un sistema doppio e cioè l'impossibilità di determinare esattamente la forma e quindi l'eccentricità dell'orbita delle componenti del sistema a partire dal loro movimento apparente sulla volta celeste.

Questa ambiguità ha reso difficile all'inizio l'analisi dei dati relativi al pulsar doppio. Questa ha tuttavia mostrato sin dall'inizio una cospicua precessione relativistica del periastro entro l'ordine di grandezza previsto per un'orbita così stretta ed eccentrica. Infine si è rilevato come il sistema vada accelerando e l'orbita si restringa denunciando una perdita di energia che può essere solamente dovuta all'emissione di onde gravitazionali.

Combinando questo effetti ed altri più sottili sulle deviazioni non kepleriane si può giungere ad una determinazione molto accurata dei dati rimanenti del sistema e ad una risoluzione pressochè totale delle ambiguità classiche. Il sistema è composto da due oggetti compatti la cui massa si aggira per ambedue sulle 1.44 masse solari.

Il lavoro di Thibaud D'Amour è stato determinante nel completare la teoria, ricondurla ad una conveniente approssimazione della relatività generale e nel conferire a questi test la grande importanza che essi meritano.

Essi infatti non solo confermano ulteriormente le misure fatte entro il sistema solare ma ci danno anche la prima conferma storica, sia pure indiretta, della esistenza di onde gravitazionali, inoltre essi non sono consistenti con le numerose varianti proposte per la relatività generale. Altri sistemi doppi vengono attualmente analizzati e non è inconcepibile che tutte queste misure aprano la via ad una nuova rivoluzione osservativa entro la teoria della relatività.

Altre conferme di natura al momento più qualitativa vengono dalla spettacolare osservazione di lenti gravitazionali in cui il potente campo di una galassia deforma e sdoppia l'immagine di una sorgente lontanissima posta al di là di questa. Anche in questo caso è prevedibile che si

passi presto a risultati più precisi e stringenti dal punto di vista quantitativo.

Giungiamo infine alla cosmologia a cui la relatività generale ha dato i contributi più profondi e sconvolgenti. La concordanza tra dati osservativi e teoria non va esente da critiche ma fa certamente riflettere. La teoria del big-bang non sarebbe nata senza l'apporto della relatività ed anche le proposte più eretiche, tra cui quella dell'universo stazionario, fanno uso del linguaggio e dei concetti della relatività.

Lo sviluppo tecnologico promette di darci misure astronomiche ed astrofisiche sempre più precise e complete e si può ben sperare di avere per l'anno 2000 una mappa molto più precisa dell'universo rispetto a quella ancora piena di ombre ed incertezze che abbiamo al presente.

Non vedo tuttavia alternative al momento della relatività generale se non alle frontiere estreme della nostra conoscenza, quelle che incontriamo quando vogliamo analizzare i primissimi istanti dell'evoluzione dell'universo.

In queste circostanze occorre una teoria unificata che realizzi il sogno incompiuto di Einstein ed in cui la teoria attuale è un frammento inadeguato. Non è inconcepibile che questa teoria abbia bisogno di un numero elevato di dimensioni dello spazio tempo per essere descritta. Al momento non possediamo questa teoria né abbiamo alcun apparato che ci permetta di vedere cosa avvenne in quegli istanti e nessuno può ragionevolmente anticipare quando questo sarà possibile. Vedremo più chiaro quando sarà operante un grande radiotelescopio con una base paragonabile a quella dell'orbita terrestre, impresa costosa ma fattibile.

Sullo spostamento del perielio dei pianeti (*)

Dionigi GALLETTO

1. Numerosi e di vario genere (modifiche *ad hoc* della legge di gravitazione universale, assunzioni relative alla forma e al comportamento della massa interna del Sole nel suo moto di rotazione, ecc.) sono stati i tentativi fatti in passato (e qualcuno anche in tempi piuttosto recenti) per spiegare, operando nel campo della meccanica newtoniana, il comportamento anomalo del perielio di Mercurio e di altri pianeti ed asteroidi.

Allo stato attuale delle conoscenze si può comunque affermare che soltanto la relatività generale può fornire una spiegazione pienamente soddisfacente di tale fenomeno. Al riguardo si può però osservare — ed è appunto questo lo scopo della presente esposizione — che la suddetta spiegazione, come si vedrà, sia pure per sommi capi, si può ottenere con tecniche di confronto sostanzialmente simili a quelle a cui si fa ricorso nell'usuale trattazione relativistica, senza fare intervenire le equazioni gravitazionali. Contemporaneamente, tramite la via qui seguita, si perviene alla celebre metrica di Schwarzschild nel vuoto, per la quale si può poi constatare, come risultato finale, che in corrispondenza ad essa risultano verificate le suddette equazioni e che quindi essa costituisce una soluzione di queste ultime.

In altri termini, mentre l'usuale trattazione relativistica dell'argomento qui affrontato è fondata sulle equazioni gravitazionali, ritenute valide *a priori*, nella presente trattazione non solo non si fa ricorso ad esse ma si fornisce anzi una conferma della loro validità [1] almeno per il caso costituito dall'argomento qui esaminato.

Sull'intera questione intendo comunque ritornare più diffusamente e nei dettagli in un prossimo lavoro.

(*) Lavoro eseguito nell'ambito del G.N.F.M. del Consiglio Nazionale delle Ricerche.

(1) Si veda a proposito quanto viene osservato, con riferimento alle equazioni gravitazionali, da S. Weinberg nel § 8.3 del suo trattato *Gravitation and Cosmology, Principles and Applications of the General Theory of Relativity*, New York, 1972.

Si tenga anche conto della precisazione che al riguardo verrà fatta alla fine del n. 5.

2. La via generalmente seguita in relatività generale per spiegare lo spostamento del perielio di Mercurio (e conseguentemente quello, molto più modesto, di altri pianeti o asteroidi) consiste:

a) nello scrivere la metrica dello spazio-tempo sotto la condizione che essa sia statica e a simmetria spaziale sferica: facendo ricorso a coordinate polari e a opportune trasformazioni che sostanzialmente si limitano a modificare la coordinata temporale t e l'originario significato della coordinata r, tale metrica può essere messa nella forma

$$ds^2 = B(r)\, c^2 dt^2 - A(r)\, dr^2 - r^2 (d\theta^2 + \sin^2\theta\, d\varphi^2), \tag{2.1}$$

dove $A(r)$ e $B(r)$ sono funzioni indeterminate;

b) nel fare ricorso alle equazioni gravitazionali di Einstein per la determinazione di $A(r)$ e $B(r)$: imponendo la condizione che per $r \to \infty$ la metrica (2.1) si riduca alla metrica di Minkowski scritta in coordinate polari, ossia imponendo le condizioni

$$\lim_{r\to\infty} A(r) = \lim_{r\to\infty} B(r) = 1,$$

tramite le equazioni gravitazionali, private ovviamente del termine cosmologico, si ottiene

$$B(r) = \frac{1}{A(r)} = 1 + \frac{a}{r}, \tag{2.2}$$

con a costante;

c) nel fare ricorso all'approssimazione newtoniana per la determinazione di a, e cioè nell'osservare, tramite il confronto con il caso newtoniano, che a grande distanza dalla massa centrale M (la massa che è all'origine della simmetria spaziale sferica e che dà luogo alla metrica (2.1)) $B(r)$ viene a identificarsi con $1 - \dfrac{2}{c^2} U(r)$, dove $U(r)$ è il potenziale newtoniano originato da M: ciò comporta, ricordando (2.2) e indicando con G la costante di gravitazione universale:

$$B(r) = \frac{1}{A(r)} = 1 - 2\frac{GM}{c^2 r}, \tag{2.3}$$

con che (2.1) diventa

$$(2.4) \qquad ds^2 = \left(1 - 2\frac{GM}{c^2 r}\right) c^2 dt^2 - \frac{1}{1 - 2\dfrac{GM}{c^2 r}} dr^2 - r^2(d\theta^2 + \sin^2\theta\, d\varphi^2),$$

che costituisce la celebre *soluzione di Schwarzschild*;

d) nello scrivere le equazioni delle geodetiche relative alla metrica (2.4), ricordando al riguardo che le equazioni delle geodetiche dello spazio-tempo (ovviamente a $ds^2 \neq 0$) descrivono il moto di una particella test (pianeta) liberamente gravitante: assumendo s come parametro si ottiene (osservando che, stante la simmetria sferica, il moto del pianeta risulta piano e che quindi si può assumere $\theta = \frac{\pi}{2}$):

$$(2.5) \qquad r^2 \frac{d\varphi}{ds} = \frac{C}{c},$$

$$(2.6) \qquad \frac{dt}{ds} = \frac{\gamma}{c^2\left(1 - 2\dfrac{GM}{c^2 r}\right)},$$

con C e γ costanti. Da (2.4), tenendo conto di (2.5), (2.6), si ottiene poi, con alcuni passaggi che per brevità non vengono qui riassunti, l'equazione differenziale che permette, tramite il suo studio, di fornire la descrizione e spiegazione del comportamento del perielio dei pianeti, in particolare di quello di Mercurio.

Osservazione. Alcuni autori applicano d) dopo a) e b), ossia calcolano le geodetiche relative alla metrica (2.1) tenendo conto di (2.2). La determinazione della costante a, e cioè delle espressioni (2.3), viene effettuata tramite il confronto dell'equazione differenziale a cui si è accennato alla fine di d) con l'analoga che si ottiene nel caso newtoniano.

3. Assumendo come punto di partenza la metrica (2.1) e s come parametro, le equazioni delle geodetiche relative a tale metrica implicano, con le condizioni iniziali $\theta(s_0) = \frac{\pi}{2}$, $\left(\frac{d\theta}{ds}\right)_{s=s_0} = 0$, che il moto si svolga nel piano $\theta = \frac{\pi}{2}$ e che risulti

$$(3.1) \qquad r^2 \frac{d\varphi}{ds} = \frac{C}{c},$$

$$(3.2) \qquad \frac{dt}{ds} = \frac{\gamma}{c^2 B(r)},$$

con C e γ costanti.

Da (2.1) segue poi

$$c^2 B(r) \left(\frac{dt}{ds}\right)^2 - A(r) \left(\frac{dr}{ds}\right)^2 - r^2 \left(\frac{d\varphi}{ds}\right)^2 = 1,$$

da cui, tenendo conto di (3.1), (3.2) e di

$$\frac{dr}{ds} = \frac{dr}{d\varphi} \frac{d\varphi}{ds} = -\frac{C}{c} \frac{d}{d\varphi} \frac{1}{r},$$

si ottiene

$$(3.3) \qquad \frac{C^2}{c^2} \left(A(r) B(r) \left(\frac{d}{d\varphi} \frac{1}{r}\right)^2 + \frac{B(r)}{r^2}\right) + B(r) = \frac{\gamma^2}{c^2}.$$

Posto

$$(3.4) \qquad B(r) = 1 + \beta(r),$$

la condizione che per $r \to \infty$ la metrica (2.1) si riduca alla metrica di Minkowski in coordinate polari comporta

$$(3.5) \qquad \lim_{r \to \infty} \beta(r) = 0,$$

mentre (3.3) si può scrivere

$$(3.6) \qquad \frac{C^2}{2}\left(A(r)\,B(r)\left(\frac{d}{d\varphi}\frac{1}{r}\right)^2+\frac{1}{r^2}\right)+\frac{c^2}{2}\beta(r)+$$

$$+\frac{C^2}{2}\,\frac{\beta(r)}{r^2}=\frac{\gamma^2-c^2}{2}.$$

4. Tenendo anche conto che le velocità dei pianeti sono piccole rispetto alla velocità della luce, si può a questo punto far ricorso all'approssimazione newtoniana, procedendo pertanto come al punto c) del n. 2, il che, come è stato in tale punto richiamato, comporta, senza far ricorso alle equazioni gravitazionali di Einstein [2]:

$$(4.1) \qquad B(r)=1-2\,\frac{GM}{c^2 r}$$

e cioè, tenendo presente (3.4):

$$(4.2) \qquad \beta(r)=-2\,\frac{GM}{c^2 r},$$

che, come è ovvio debba essere, è in accordo con (3.5).

L'espressione (4.2) comporta che l'ultimo addendo a primo membro nell'equazione (3.6) si espliciti in

$$\frac{C^2}{2}\,\frac{\beta(r)}{r^2}=-\frac{C^2}{c^2}\,\frac{GM}{r^3}$$

e sia pertanto, a grandi distanze, trascurabile rispetto al termine $\frac{c^2}{2}\beta(r)$, con che l'equazione (3.6), tenendo presente (4.2), si riduce a

(2) Meglio ancora, come verrà esposto nel lavoro a cui ho accennato al n. 1, all'espressione (4.1) si può pervenire attraverso considerazioni dirette e opportuni calcoli che hanno come punto di partenza l'espressione (3.3) e l'esame del caso in cui le orbite siano circolari. Tale via tra l'altro non comporta la condizione che il confronto con il caso newtoniano debba avvenire a «grande distanza» da M. Per motivi di brevità la suddetta via non viene per ora esposta.

$$(4.3) \qquad \frac{C^2}{2}\left(A(r)\left(1-2\frac{GM}{c^2 r}\right)\left(\frac{d}{d\varphi}\frac{1}{r}\right)^2+\frac{1}{r^2}\right)-\frac{GM}{r}=E,$$

dove si è fatta la posizione

$$\frac{\gamma^2-c^2}{2}=E.$$

Procedendo nel ricorso all'approssimazione newtoniana si imponga la condizione, in accordo con quanto la teoria newtoniana comporta, che le orbite siano ellittiche. Ciò comporta che nell'espressione (4.3) risulti

$$A(r)\left(1-2\frac{GM}{c^2 r}\right)=\text{costante},$$

da cui segue [(3)], tenendo presente (4.1) e ricordando che deve essere $\lim_{r\to\infty} A(r)=1$:

$$(4.4) \qquad A(r)=\frac{1}{B(r)}=\left(1-2\frac{GM}{c^2 r}\right)^{-1},$$

ossia (2.3).

In sintesi, si può quindi concludere che la determinazione esplicita delle funzioni $A(r)$ e $B(r)$ che compaiono nella metrica (2.1), e quindi la determinazione esplicita della metrica di Schwarzschild (2.4), si può effettuare, almeno per grandi distanze dalla massa centrale M, facendo ricorso a metodi di approssimazione e di confronto che, anche se non gli stessi, sono nella sostanza simili a quelli a cui si fa ricorso nell'usuale trattazione relativistica, senza dover ricorrere alle equazioni gravitazionali di Einstein [(4)].

Poiché, come è stato brevemente richiamato al n. 2, è proprio operando con la suddetta metrica che si perviene alla spiegazione del com-

---

(3) Si veda al riguardo la nota (4).

(4) Sulla deduzione della metrica (2.4) qui presentata mi riprometto, come già ho accennato al n. 1, di ritornare più diffusamente e a fondo nel lavoro a cui accenno anche nella nota 2, soprattutto per quanto concerne la deduzione delle espressioni (4.4).

portamento del perielio di Mercurio e di altri pianeti (i quali si possono considerare tutti a grande distanza dal Sole), si può pertanto affermare che tale spiegazione si può effettuare senza dover ricorrere alle equazioni gravitazionali di Einstein, le quali, tra l'altro, comportano calcoli alquanto laboriosi.

La sostituzione di (4.1), (4.2), (4.4) nell'equazione (3.6) dà infatti luogo all'equazione

$$\left(\frac{d}{d\varphi}\frac{1}{r}\right)^2+\frac{1}{r^2}-2\frac{GM}{C^2}\frac{1}{r}-2\frac{GM}{c^2}\frac{1}{r^3}=\frac{\gamma^2-c^2}{C^2},$$

da cui, derivando rispetto a $\frac{1}{r}$, si ottiene

$$\frac{d^2}{d\varphi^2}\frac{1}{r}+\frac{1}{r}=\frac{GM}{C^2}+3\frac{GM}{c^2}\frac{1}{r^2}, \tag{4.5}$$

che è la forma esplicita dell'equazione a cui si è fatto cenno al punto d), la quale permette, tramite il suo studio, di fornire la descrizione e spiegazione del comportamento del perielio dei pianeti, in particolare di quello di Mercurio. Essa differisce dall'analoga che si ottiene nel caso newtoniano per il termine aggiuntivo $3\,\frac{GM}{c^2}\frac{1}{r^2}$.

L'equazione (4.5) pur potendosi ottenere, come si è visto, sia pure per sommi capi, senza far ricorso alle equazioni gravitazionali, resta fermo il fatto che nel dedurre tale equazione si è fatto implicitamente ricorso a quello che può essere considerato come il principio che è alla base di tutta la teoria della relatività generale, e cioè che ogni massa (il Sole in questo caso) influenzi (nel presente caso, in cui si ha soltanto il campo gravitazionale generato da M, determini) la geometria dello spazio-tempo e che ogni «particella test» (i pianeti e gli asteroidi in questo caso) segua una geodetica di questo, sotto l'azione della sola curvatura. Alla luce di quanto esposto, il comportamento del perielio dei pianeti va quindi intanto visto come un grande test, una grande verifica di tale principio.

5. Indicando con R_{ij} il tensore di curvatura contratto, in corrispondenza alla metrica (2.4) (ritenuta ovviamente dedotta tramite la via qui indicata) si ottiene per tale tensore:

$$R_{ij} = 0,$$

che è quanto basta per poter affermare che in corrispondenza a tale metrica risultano verificate le equazioni gravitazionali e che quindi essa costituisce una soluzione di queste.

Si può perciò concludere che nella presente trattazione non solo non si fa ricorso alle equazioni gravitazionali ma (una volta imposta la condizione che, trascurando, per $r \to \infty$, gli infinitesimi di ordine superiore al primo rispetto all'infinito principale $\beta(r)$ espresso da (4.2), l'equazione (3.6) si riduca, come la sua espressione suggerisce, all'integrale dell'energia del caso newtoniano) addirittura si prova la loro validità, almeno per il caso costituito dall'argomento qui esaminato.

L'Autore ringrazia il Prof. B. Barberis per la collaborazione datagli nella stesura del presente lavoro.

L'attività di Giuseppe Colombo, nel campo della meccanica non lineare

Dario GRAFFI

1. Giuseppe Colombo, come ognuno ben sa, si è occupato, dal 1962 fino alla Sua prematura scomparsa, di Astronautica e di Meccanica Celeste, portando a questi argomenti contributi di fondamentale importanza. Ma la Sua opera scientifica non si è limitata agli argomenti ora citati; fra l'altro si è occupato di Meccanica non-lineare (o meglio di oscillazioni non-lineari); e precisamente, negli anni cinquanta, quando questo ramo della Fisica-matematica era di viva attualità.

Ritengo opportuno riassumere in questa comunicazione i di Lui contributi alla Meccanica non-lineare, anche perché essi già dimostrano la Sua preparazione matematica e fisica, la Sua abilità nell'applicazione della Matematica a problemi concreti (fra l'altro, in quell'epoca, era forse l'unico cultore di Meccanica Razionale che conoscesse e sapesse applicare la Topologia, così importante per gli sviluppi teorici della Meccanica non-lineare); e la capacità di immaginare opportuni modelli, in ispecie meccanici, di importanti fenomeni non lineari; abilità ed immaginazione che si ritrovano nella Sua grandiosa opera successiva, alla quale è tanto giustamente dedicato l'odierno Simposio.

Con questa comunicazione intendo anche tributare un omaggio alla Memoria di un Amico che tanto abbiamo amato ed ammirato.

2. Come ho accennato or ora, e come si vede dalla Bibliografia qui allegata, le prime Memorie che Giuseppe Colombo ha dedicato alla Meccanica non-lineare risalgono al 1950. A quell'epoca Egli si era già aggiornato sui risultati ottenuti nella Meccanica non-lineare: risultati che consistevano essenzialmente nell'opera pionieristica di B. Van der Pol, nell'opera della Scuola russa, in particolare di N. Minorsky (al quale spetta, fra l'altro, il merito di aver diffuso in Occidente i lavori di quella Scuola); nei contributi di molti altri studiosi, fra i quali ci limitiamo qui a citare N. Levinson, J.E. Littlewood, L.M. Cartwright e, fra gli italiani, il compianto Giovanni Sansone. Tutte queste ricerche erano però

dedicate in massima parte a sistemi meccanici od elettrici ad un solo grado di libertà (nei quali si rivelano spesso preziosi i metodi di Poincarè-Bendixon) spesso svolte con metodi approssimati, quindi in condizioni molto particolari.

Ora, già nei Suoi primi lavori, Colombo si occupa di equazioni o sistemi non lineari di ordine superiore al secondo; e di tali questioni si occuperà spesso in seguito. Non mancano però, nella Sua Opera, lavori relativi a sistemi ad un grado di libertà; Egli considera però generalmente problemi ai quali non è lecito applicare metodi approssimati; oppure istituisce metodi di approssimazione sostanzialmente nuovi, atti ad affrontare nuovi problemi.

3. In due note che risalgono al 1950 (vedi Bibl.: 1950/1; 1950/2) e in una conferenza tenuta nel 1954, a Varenna, nel primo corso C.I.M.E. (1954/2), Colombo considera un problema elettrico che si rieonduce ad una equazione non-lineare del terzo ordine di cui dimostra l'esistenza di una soluzione periodica; e dopo poco tempo si occupa ancora dell'esistenza di soluzioni periodiche, ma per un sistema a due gradi di libertà, riferendosi, come modello, ad un sistema elettrico che ora descriveremo. Consideriamo due circuiti di Van der Pol, ognuno, come è noto, formato da una induttanza, una capacità ed una resistenza, in serie con un dipolo non lineare ad N; i circuiti sono accoppiati tra loro mediante una capacità (fig. 1).

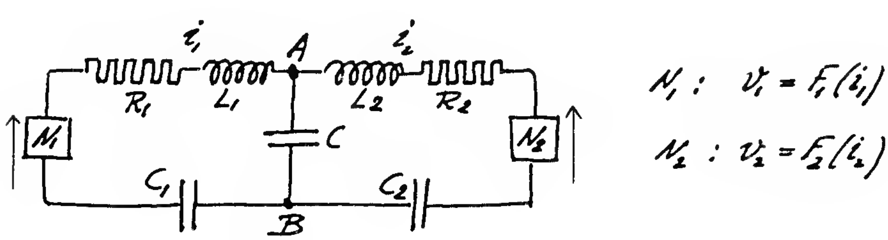


Figura 1

Siano R_1, L_1, C_1, N_1; R_2, L_2, C_2, N_2, le resistenze, le induttanze, le capacità e i dipoli non lineari, rispettivamente del primo e del secondo circuito orientati questi ultimi come in fig. 1. I due circuiti sono accoppiati tramite la capacità C. Siano q_1 e q_2 le cariche sulle armaturc positive di C_1, C_2 rispettivamente; o meglio sulle prime armature incontrate

percorrendo i circuiti nei loro versi positivi; i_1, i_2 le correnti; v_1, v_2 le differenze di potenziale agli estremi di N_1 e N_2; V_A e V_B i potenziali delle armature di C.

Valgono le equazioni:

$$(3.1)_1 \qquad L_1 \frac{di_1}{dt} + R_1 i_1 + \frac{q_1}{C_1} + V_A - V_B + v_1 = 0$$

$$(3.1)_2 \qquad L_2 \frac{di_2}{dt} + R_2 i_2 + \frac{q_2}{C_2} + V_A - V_B + v_2 = 0$$

con

$$(3.2) \qquad i_1 = \frac{dq_1}{dt} \;, \; i_2 = \frac{dq_2}{\mathrm{dt}} \;, \; \frac{d}{dt}(V_A - V_B) = \frac{1}{C}(i_1 + i_2).$$

Inoltre, si ammette che sia:

$$(3.3) \qquad v_1 = F_1(i_1) \;, \; \frac{dv_1}{dt} = F_1'(i_1) \frac{di_1}{dt}$$

e analoga formula per v_2.

Allora, derivando rispetto al tempo $(3.1)_1$ e $(3.1)_2$ e tenendo conto delle (3.3) si ha

$$(3.4) \qquad L_1 \frac{d^2 i_1}{dt^2} + (R_1 + F'(i_1)) \frac{di_1}{dt} + \left(\frac{1}{C} + \frac{1}{C_1}\right) i_1 + \frac{i_2}{C} = 0$$

$$(3.5) \qquad L_2 \frac{d^2 i_2}{dt^2} + (R_2 + F'\;(i_2)) \frac{di_2}{dt} + \left(\frac{1}{C} + \frac{1}{C_2}\right) + \frac{i_l}{C} = 0.$$

Ora, nel caso di Van der Pol si può scrivere:

$$(3.6) \qquad \frac{R_1 + F'(i_1)}{L_1} = -(\alpha_1 - \beta_1 i_1^2), \; \frac{R_2 + F'(i_2)}{L_2} = -(\alpha_2 - \beta_2\, i_2^2)$$

(α_1, α_2, β_1, β_2 positive); allora posto:

$$i_1 = x,\ i_2 = y,\ \omega_1^2 = \frac{1}{L_1(C_1 + C)},\ \omega_2^2 = \frac{1}{L_2(C_2 + C)},$$

$$m_1 = \frac{1}{L_1 C},\ m_2 = \frac{1}{L_2 C},$$

dopo aver diviso (3.4) per L_1, (3.5) per L_2, si ha:

$$(3.6)_1 \qquad \frac{d^2x}{dt^2} - (\alpha_1 - \beta_1 x^2)\,\dot{x} + \omega_1^2\, x + m_1\, y = 0$$

$$(3.6)_2 \qquad \frac{d^2y}{dt^2} - (\alpha_2 - \beta_2 y^2)\,\dot{y} + \omega_2^2\, y + m_2\, x = 0 \ (*)$$

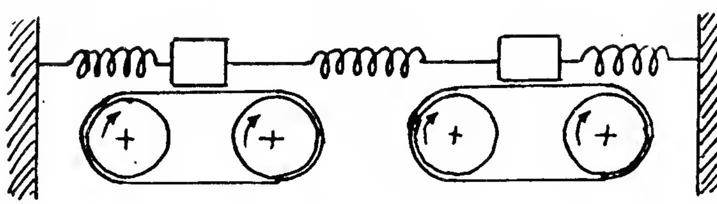


Figura 1 bis

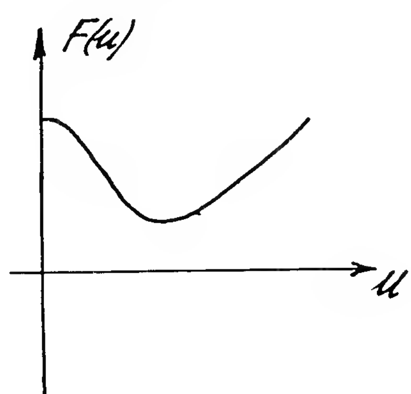


Figura 1 ter

(*) Un interessante sistema meccanico studiato da G. Colombo in (1952/1) è quello schematizzato in fig. 1 bis, quando si ammetta che la forza di attrito dinamico $F(u)$ dipenda dalla velocità relativa u secondo una legge del tipo indicato in fig. 1 ter.

Ora Colombo ricava, nella (1950/2) come già si è detto, l'esistenza di soluzioni periodiche per il sistema (3.6), almeno supponendo m_1, m_2 sufficientemente piccoli.

Riportiamo qui, per sommi capi, la Sua dimostrazione:

posto $\dot{x}=u$, $\dot{y}=v$, lo spazio delle fasi x, u, y,v è lo spazio quadrimensionale S_4. In S_4 considera due piani bidimensionali, $\pi_1(x, u)$, $\pi_2(y, v)$ e dimostra che in ambedue detti piani esistono due linee chiuse (F_1, F_1'), (F_2, F_2'), con F_1', F_2', rispettivamente, interne ad F_1 ed F_2, e con l'origine interna a F_1', F_2'; quindi (F_1, F_1'), (F_2, F_2') determinano due regioni anulari: R_1 in π_1, R_2 in π_2 equivalenti, topologicamente, ciascuna ad una corona circolare (fig. 2).

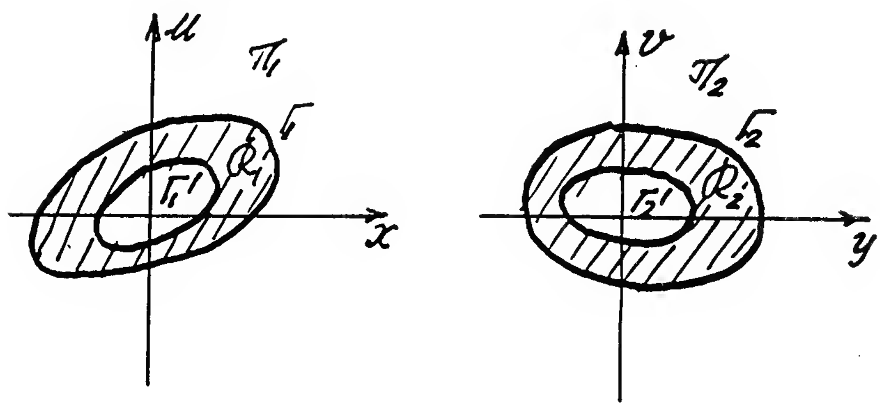


Figura 2

Ora Colombo prova che se:

$$x=x\ (t),\ u=u\ (t),\ y=y\ (t),\ v=v\ (t) \tag{3.7}$$

è una linea di S_4 con x (0), u (0) interna a R_1; e con y (0), v (0) interna ad R_2, la linea rappresentata da x (t), y (t) è per $t > 0$ interna ad R_1; mentre la linea y (t), v (t) è per $t > 0$ interna ad R_2.

Si intuisce (ma la dimostrazione che di ciò dà Colombo non è affatto semplice) che esiste in S_4 una linea c di equazioni:

$$x^*=x^*\ (t),\ u^*=u^*\ (t),\ y^*=y^*\ (t),\ v^*=v^*\ (t) \tag{3.8}$$

tale che le linee x^* (t), u^* (t), y^* (t), v^* (t), siano rispettivamente chiuse in R_1, R_2. Allora la c di S_4, è una linea chiusa, che può rappresentare

una soluzione periodica del sistema (3.6).

Colombo non dimostra la stabilità di questa soluzione nel senso classico; ma dimostra che ogni soluzione di (3.6) prossima alla soluzione periodica rimane sempre in un intorno limitato della soluzione periodica stessa.

A questa ricerca si può collegare la nota (1955/2) nella quale Colombo considera il sistema:

$$(3.9)_1 \qquad \ddot{x} + f_1(x, \dot{x}) + g_1(x) = \epsilon F_1(x, \dot{x}, y, \dot{y})$$

$$(3.9)_2 \qquad \ddot{y} + f_2(y, \dot{y}) + g_2(y) = \epsilon F_2(x, \dot{x}; y, \dot{y})$$

che in certo senso generalizza (3.6) e dove ϵ è un parametro positivo o nullo. Ora, se è $\epsilon = 0$ le due equazioni di (3.9) si separano, cioè ognuna rappresenta un sistema ad un grado di libertà. In questo caso Colombo ammette che tanto $(3.9)_1$ quanto $(3.9)_2$ abbiano una soluzione periodica rappresentata rispettivamente in π_1, e π_2 da un ciclo limite stabile. Più precisamente, ammette che queste soluzioni di $(3.9)_1$ e $(3.9)_2$ siano rappresentate dalle relazioni:

$$(3.10) \qquad x = x(t), \; u = u(t); \; y = y(t + \varphi), \; v = v(t + \varphi)$$

(dopo una ovvia ed opportuna scelta dell'origine e dell'unità del tempo).

Le prime due funzioni si ammettono periodiche di periodo 2π; le altre due, di periodo $2\pi\lambda$. Ad ogni valore del parametro φ corrisponde nello spazio S_4 una linea chiusa e al variare di φ fra 0 e $2\pi\lambda$ queste linee coprono in S_4 una varietà chiusa bidimensionale V_0.

Se è invece $\epsilon \neq 0$, ma sufficientemente piccolo, Colombo dimostra, con profonde considerazioni analitiche, che esiste una varietà chiusa V^*, ricoperta da linee soluzioni del sistema (3.9) che rappresentano moti, in generale, aperiodici. Poiché per condizioni iniziali prossime a V^*, le soluzioni di (3.9) tendono per $t \to \infty$ alle linee di V^* i moti presentano carattere di stabilità.

4. Nella nota (1952/1) Colombo riprende il sistema (3.6); e, ammettendo alcune ipotesi semplificatrici, riesce ad ottenere importanti proprietà di alcune soluzioni del sistema stesso. Egli ammette anzitutto deboli tanto la non linearità quanto l'accoppiamento fra le due equazioni; ammette inoltre che i periodi delle soluzioni qualora le due equazioni di

(3.6) siano disaccoppiate (cioè per $m_1 = m_2 = 0$) e manchi la non linearità ($\alpha = \beta = 0$) siano poco diversi. Poiché con opportune scelte della unità di misura del tempo si può assumere $\omega_1 = 1$ risulterà, per le predette ipotesi, che ω_2 è poco diversa da 1. Per esprimere matematicamente queste ipotesi converrà al solito introdurre il parametro ϵ sufficientemente piccolo e tale da poter fare alcune altre ipotesi, sulle quali qui non insisteremo; e scrivere infine il sistema (3.6) nella seguente forma:

$$(4.1)_1 \qquad \ddot{x} + x = \epsilon \, [(\alpha_1 - \beta_1 x^2)\, \dot{x} - m_1 y]$$

$$(4.1)_2 \qquad \ddot{y} + y = \epsilon \, [(\alpha_2 - \beta_2 y^2)\, \dot{y} - (\varrho y + m_2 x)]$$

dove $\epsilon\varrho$ rappresenta il divario tra ω_1 e ω_2. Per determinare se il sistema (4.1) ammette soluzioni periodiche approssimate con approssimazione dell'ordine di ϵ, Colombo procede applicando un metodo proposto dalla Cartwright. Noi perverremo agli stessi risultati applicando il metodo classico di Kryloff-Bogoliuboff (che denoteremo in seguito, per brevità, metodo K.B.).

A tale proposito poniamo:

$$(4.2) \qquad \begin{aligned} &x = a \text{ sen } (t + \psi), \; \dot{x} = a \cos (t + \psi), \\ &y = h \text{ sen } (t + u), \; \dot{y} = h \cos (t + u) \end{aligned}$$

con a, ψ, h, u funzioni di t (qui, ovviamente, la u ha un significato diverso da quello del paragrafo 3.). Applicando il citato metodo K.B. si ha che:

$$(4.3) \qquad \dot{a} = \frac{\epsilon}{2\pi} \int_0^{2\pi} F_1(x, \dot{x}, y, \dot{y}) \cos (t + \psi)\, dt,$$

$$(4.4) \qquad \dot{h} = \frac{\epsilon}{2\pi} \int_0^{2\pi} F_2(x, \dot{x}, y, \dot{y}) \cos (t + u)\, dt,$$

$$(4.5) \qquad \dot{\psi} = -\frac{\epsilon}{2\pi a} \int_0^{2\pi} F_1(x, \dot{x}, y, \dot{y}) \text{ sen } (t + \psi)\, dt,$$

$$(4.6) \qquad \dot{u} = -\frac{\epsilon}{2\pi h} \int_0^{2\pi} F_2(x, \dot{x}, y, \dot{y}) \text{ sen } (t + u)\, dt,$$

dove $\epsilon F_1 (x, \dot{x}, y, \dot{y})$, $\epsilon F_2 (x, \dot{x}, y, \dot{y})$ sono rispettivamente i secondi membri di $(4.1)_1$, $(4.1)_2$. Ora, sempre secondo il metodo K.B., gli integrali (4.3), (4.4) e (4.5), (4.6) si possono calcolare ponendo le (4.2) al posto di x, $\dot{x}$, y, $\dot{y}$ e trattando a, h, ψ, u come costanti. Si trova, dopo alcuni calcoli:

$$\dot{a} = \frac{\epsilon}{2}\left[\alpha_1 a \left(1 - \frac{a^2}{a_0^2}\right) - m_1 h \operatorname{sen}(u-\psi)\right] \tag{4.7}$$

$$\dot{h} = \frac{\epsilon}{2}\left[\alpha_2 h \left(1 - \frac{h^2}{h_0^2}\right) + m_2 a \operatorname{sen}(u-\psi)\right] \tag{4.8}$$

$$\dot{\psi} = \frac{\epsilon}{2a} h\, m_1 \cos(u-\psi) \tag{4.9}$$

$$\dot{u} = \frac{\epsilon}{2h} a\, m_2 \cos(u-\psi) + \frac{\epsilon}{2}\varrho \tag{4.10}$$

con $a_0^2 = \dfrac{4\,\alpha_1}{\beta_1}$; $h_0^2 = \dfrac{4\,\alpha_2}{\beta_2}$.

Poniamo ora:

$$\varphi = u - \psi \tag{4.11}$$

e cerchiamo sotto quali condizioni la grandezza φ è costante ossia sotto quali condizioni è $\dot{u} - \dot{\psi} = 0$. Si ha allora l'equazione:

$$\left(\frac{a}{h} m_2 - \frac{h}{a} m_1\right) \cos\varphi + \varrho = 0. \tag{4.12}$$

Cerchiamo ora i valori di a e h di regime per le $(4.1)_1$, $(4.1)_2$. In corrispondenza di questi particolari di a e h debbono essere nulli i secondi membri di (4.7) e (4.8): si hanno così le equazioni

$$\alpha_1 a \left(1 - \frac{a^2}{a_0^2}\right) - m_1 h \operatorname{sen}\varphi = 0 \tag{4.13}$$

$$\alpha_2 h \left(1 - \frac{h^2}{h_0^2}\right) + m_2 a \operatorname{sen} \varphi = 0 \ . \tag{4.14}$$

Le (4.12), (4.13), (4.14) costituiscono un sistema da cui si possono determinare eventuali valori di regime di a, h, φ.

Va notato che in questo caso si ha, dalle (4.9) (4.10) e tenendo presente (4.12):

$$\psi = \frac{\epsilon\, m_1 h}{2a}\, t \cos \varphi + \psi_0 \ , u = \frac{\epsilon\, m_1 h}{2a}\, t \cos \varphi + u_0 . \tag{4.15}$$

Quindi:

$$\begin{aligned} x &= a \operatorname{sen} \left[t \left(1 + \frac{\epsilon\, m_1\, h \cos \varphi}{2a}\right) + \psi_0 \right] , \\ u &= h \operatorname{sen} \left[t \left(1 + \frac{\epsilon\, m_1 h \cos \varphi}{2a}\right) + u_0 \right] . \end{aligned} \tag{4.16}$$

Le (4.16) rappresentano, in corrispondenza ad ogni soluzione del sistema (4.12), (4.13), (4.14) una soluzione periodica di $(4.1)_1$ e $(4.1)_2$, con periodo:

$$T = \frac{2\pi}{1 + \dfrac{\epsilon\, m_1 h}{2\,a} \cos \varphi} = 2\pi \left[1 - \frac{\epsilon\, m_1 h}{a} \cos \varphi\right] + O\,(\epsilon^2) \ .$$

E ciò, conforme al risultato ottenuto da Colombo, che risolve anche il sistema ora considerato con eleganti considerazioni geometriche ed analitiche. Egli determina infine, almeno in casi particolari, le condizioni di stabilità per il moto rappresentato da (4.16). Non esporremo qui questo Suo risultato, limitandoci a segnalare che in certi casi esistono per il sistema $(4.1)_1$, $(4.1)_2$ tre soluzioni approssimate periodiche stabili. D'altra parte la stessa questione sarà approfondita nel caso particolare che costituisce oggetto del seguente paragrafo.

5. Ci proponiamo di applicare qui i risultati del paragrafo precedente al seguente problema trattato da Colombo in (1952/2). In questo lavoro, Egli considera dapprima un sistema elettrico (fig. 3a) e immagina un sistema meccanico (cfr. fig. 3b) (*) che conduce ad equazioni del tipo $(4.1)_1$, $(4.1)_2$. Egli perviene così alle equazioni:

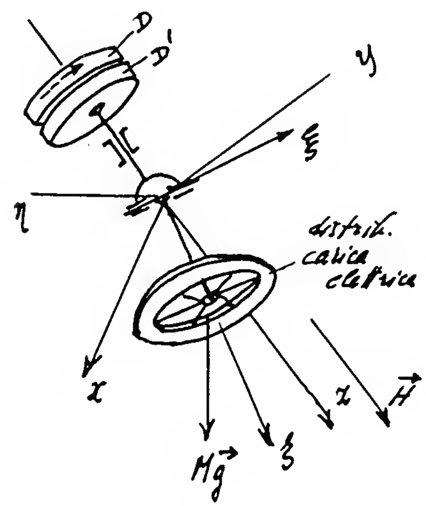


Figura 3b

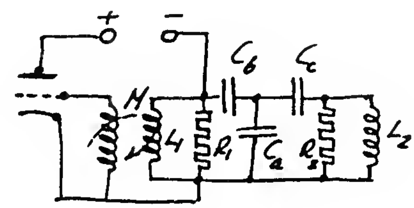


Figura 3a

$$(5.1)_1 \qquad \ddot{x}+x=\epsilon\,(my+(\alpha-\beta\,x^2)\,\dot{x})$$

$$(5.1)_2 \qquad \ddot{y}+y=\epsilon\,(nx+b\dot{x}+\varrho y)$$

che coincidono con $(4.1)_1$, $(4.1)_2$ a condizione che si ponga α, β in luogo di α_1, β_1; m ed n in luogo di $-m_1$, $-m_2$; $b\dot{x}$ in luogo di $(\alpha_2-\beta_2 y^2)\,\dot{y}$ e ϱ in luogo di $-\varrho$.

(*) Attorno all'asse z (non verticale) di un riferimento cartesiano fisso Ωxyz avente l'asse x orizzontale, può ruotare un corpo rigido D' solidale con la forcella di un giunto cardanico; attorno all'asse ξ della forcella (asse ovviamente appartenente al piano xy) è girevole un secondo corpo rigido ϑ, pesante, avente il proprio baricentro su un'asse ζ normale a ξ e non coincidente con l'origine Ω. Il corpo ϑ è carico (cioè sede di una distribuzione di cariche elettriche); ed il suo elissoide d'inerzia rispetto al punto Ω è di rotazione rispetto all'asse ζ. Sul corpo ϑ agisce inoltre un campo magnetico uniforme e costante di intensità H.

Il corpo rigido D' è trascinato in rotazione dalla coppia (non lineare) di attrito dinamico che si manifesta al contatto fra il disco D' stesso con un altro disco D, mantenuto in rotazione con velocità angolare costante.

Risolvendo il sistema con le posizioni (4.2) si trovano ancora le (4.8), (4.9) salvo l'accennato scambio dei simboli. Si ha cioè

$$\dot{a} = \frac{\epsilon}{2}\left[\alpha\, a\left(1 - \frac{a^2}{a_0^2}\right) + mh \text{ sen } \varphi\right] \tag{5.2}$$

$$\dot{\psi} = -\epsilon\, m\, \frac{h}{2a} \cos \varphi \tag{5.3}$$

$$\begin{aligned}\dot{h} &= \frac{\epsilon}{2\pi h} \int_0^{2\pi} (b\dot{x} + \varrho\, y + nx) \cos (t+u)\, dt = \\ &= \frac{\epsilon}{2}\left(\frac{a}{h}\, b \cos (u-\psi) - n\, \frac{a}{h} \text{ sen } (u-\psi)\right)\end{aligned} \tag{5.4}$$

$$\begin{aligned}\dot{u} &= -\frac{\epsilon}{2\pi\, h} \int_0^{2\pi} (b\,\dot{x} + \varrho\, y + nx) \text{ sen } (t+u)\, dt = \\ &= -\frac{\epsilon}{2\, h}\left(ba \text{ sen } (u-\psi) + \varrho + na \cos (u-\psi)\right)\end{aligned} \tag{5.5}$$

Se ora si pone, in accordo con quanto detto nel paragrafo 4., $u - \psi = \varphi$ con φ costante, si ha l'equazione:

$$-b\, \frac{a}{h} \text{ sen } \varphi - n\, \frac{a}{h} \cos \varphi + m\, \frac{h}{a} \cos \varphi - \varrho = 0. \tag{5.6}$$

In condizioni di regime, poiché $\dot{a}$ e $\dot{h}$ sono nulle si ottengono le equazioni da associare a (5.6):

$$\alpha\, a\left(1 - \frac{a^2}{a_0^2}\right) + m\, h \text{ sen } \varphi = 0 \tag{5.7}$$

$$a\, (b \cos \varphi - n \text{ sen } \varphi) = 0. \tag{5.8}$$

Poiché, come vedremo, è $a \neq 0$, dalla (5.8) si ha

$$(5.9) \qquad \operatorname{tang} \varphi = \frac{b}{n} \ ;$$

segue da ciò:

$$(5.10) \qquad b \operatorname{sen} \varphi + n \cos \varphi =$$

$$= n (\operatorname{tang} \varphi \operatorname{sen} \varphi + \cos \varphi) = \frac{n}{\cos \varphi} \ .$$

Moltiplicando (5.6) per $\frac{h}{a}$ e tenendo conto di (5.10) si ha

$$(5.11) \qquad m \cos \varphi \left(\frac{h}{a} \right)^2 - \varrho \frac{h}{a} - \frac{n}{\cos \varphi} = 0.$$

Se ora poniamo

$$(5.12) \qquad \gamma = \varrho \pm \sqrt{\varrho^2 + 4\, mn} \ ,$$

$$\gamma_+ = \varrho + \sqrt{\varrho^2 + 4\, mn}$$

$$\gamma_- = \varrho - \sqrt{\varrho^2 + 4\, mn}$$

si ha

$$(5.13) \qquad \frac{h}{a} = \frac{\gamma}{2\, m \cos \varphi} \ .$$

Poiché γ ha due valori e, come vedremo, noto $\frac{h}{a}$ potremo calcolare, da (5.7) le grandezze a, h, concludiamo che esistono due soluzioni periodiche approssimate del sistema $(5.1)_1$, $(5.1)_2$; soluzioni che indicheremo rispettivamente con s_- e s_+ a seconda che nella (5.13) si ponga in luogo di γ, γ_- o γ_+ rispettivamente. Si noti che essendo $\gamma_- \gamma_+ = -4\, mn$, i valori di $\frac{h}{a}$ in corrispondenza di s_+ e di s_- sono

$$(5.14) \qquad \left(\frac{h}{a}\right)_+ = \frac{\gamma_+}{2\,m\cos\varphi} = -\,\frac{2\,n}{\gamma_-\cos\varphi}\;;$$

$$\left(\frac{h}{a}\right)_- = -\,\frac{2\,n}{\gamma_+\cos\varphi}\,.$$

Calcoliamo ora a in condizioni di regime cioè quando si ha $\dot{a}=0$. Si ha, nel caso di s_+, ricordando anche la (5.8):

$$(5.15) \qquad \frac{a_+^2}{a_0^2} = 1 + \frac{m}{\alpha}\left(\frac{h}{a}\right)_+ \operatorname{sen}\varphi = $$

$$= 1 - \frac{2\,mn}{\alpha\gamma_-\cos\varphi}\operatorname{sen}\varphi = 1 - \frac{2\,mb}{\alpha\gamma_-}$$

e, in modo analogo si trova:

$$(5.16) \qquad \frac{a_-^2}{a_0^2} = 1 - \frac{2mb}{\alpha\gamma_+}\,.$$

Da queste equazioni si ricavano subito a_+ e a_- e da (5.14), h_+ e h_-.
Ora, affinché a_+ ed a_- (e conseguentemente anche h_+ e h_- siano reali, deve essere rispettivamente

$$(5.17)_1 \qquad \frac{2mb}{\alpha\gamma_-} < 1\;,$$

$$(5.17)_2 \qquad \frac{2mb}{\alpha\gamma_+} < 1\;.$$

Tenendo conto del fatto che γ_+ è positivo la $(5.17)_2$ è soddisfatta per $b < \dfrac{\alpha\gamma_+}{2\,m}$.

Tenendo conto che γ_- è invece negativa, la $(5.17)_1$ è verificata soltanto se $b > \dfrac{\alpha\gamma_-}{2\,m}$.

In conclusione, se è $b > \dfrac{\alpha\gamma_+}{2\,m}$ esiste soltanto s_+; per $\dfrac{\alpha\gamma_-}{2\,m} <$ $< b < \dfrac{\alpha\gamma_+}{2\,m}$ esistono entrambe le due soluzioni periodiche s_-, s_+; per $b < \dfrac{\alpha\gamma_-}{2\,m}$ esiste la sola soluzione periodica s_-: e tutto ciò conforme al grafico di figura 4.

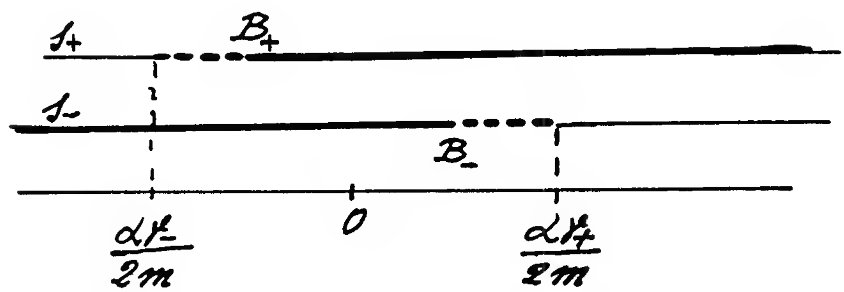


Figura 4

Si noti, nel grafico, che nel caso di s_+, il tratto $\dfrac{\alpha\gamma_-}{2\,m}$, B_+ nel caso di s_-, il tratto $\dfrac{\alpha\gamma_+}{2\,m}$, B_- sono punteggiati: ciò perché Colombo ha dimostrato, sotto alcune condizioni, che le corrispondenti soluzioni non sono stabili; quindi le soluzioni accettabili per s_+ sono quelle per cui b è nel tratto $B_+, +\infty$; quelle di s_- nel tratto $-\infty, B_-$. Scelto allora un valore di b per un sistema che oscilli su s_+, si faccia diminuire b fino a raggiungere B_+: allora la soluzione diventa instabile e subisce un salto (jump) passando su s_-. Aumentando b il sistema continua ad oscillare su s_- fino a che b raggiunge il valore corrispondente a B_-, sicché la soluzione salta su s_+. E tale rimane fino a che b è crescente. È questo il fenomeno cosiddetto di isteresi oscillatoria ben confermato dall'esperienza.

In sostanza, quindi, nel lavoro (1952/2) Colombo ha sviluppato una nuova teoria dell'isteresi oscillatoria e ne ha anche indicato un cospicuo esempio meccanico.

6. Un altro interessante fenomeno studiato da Colombo è la cosiddetta «azione asincrona» (1955/1) che si manifesta in alcuni sistemi elettromeccanici e che consiste, in sostanza, nel compensare lo smorzamento di un sistema meccanico per l'intervento indiretto di azioni elettriche. Nella memoria citata Colombo descrive un sistema meccanico che potrebbe presentare il fenomeno citato. Il dispositivo è formato (cfr. la figura 5) da due corpi Ω_1, Ω_2 vincolati a percorrere due guide rettilinee,

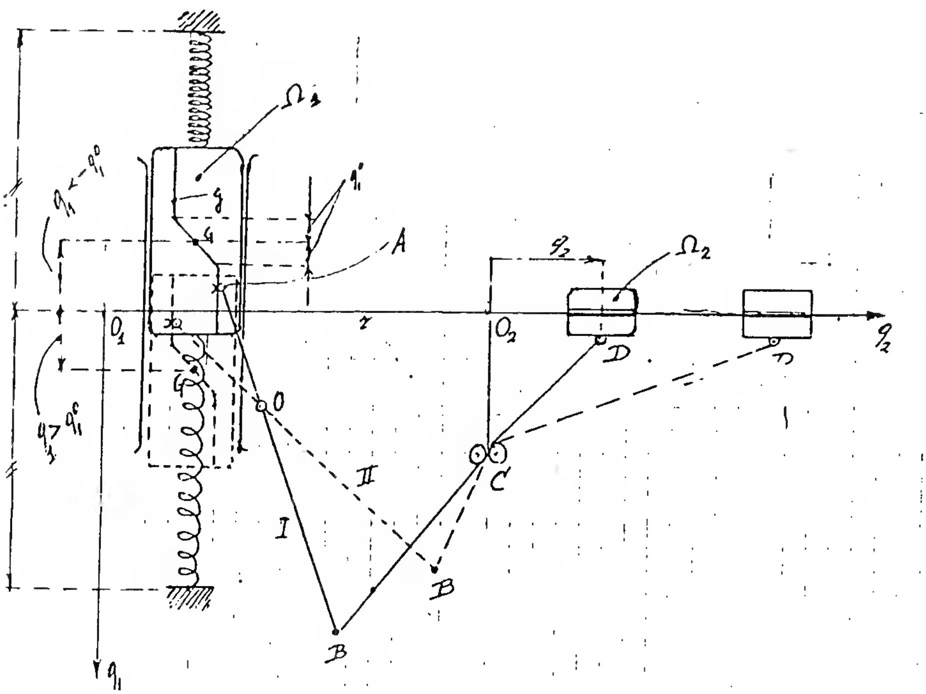


Figura 5

l'una verticale, l'altra orizzontale, sicché il sistema ha due gradi di libertà determinati dalle coordinate di Ω_1, Ω_2 o meglio da quelle dei loro baricentri rispetto alle rispettive guide. Se i due corpi non fossero vincolati fra loro le forze agenti su Ω_1 e Ω_2 sarebbero entrambe di tipo elastico e viscoso; inoltre, su Ω_2, e, rispetto al riferimento O_1 q_1 q_2 (cui si intende riferito il moto del sistema qui in esame), agisce anche una forza impressa sinusoidale E sen ωt (dovuta al fatto che tutto il sistema O_1 q_1 q_2 è animato, rispetto ad un osservatore inerziale, da moto traslatorio rettilineo armonico di pulsazione ω nella direzione O_1q_2). I due corpi sono però accoppiati mediante un dispositivo non lineare che Colombo

schematizza ammettendo che esso agisca per un tempo molto breve cioè che le sue azioni su Ω_1 possano ritenersi come impulsive ed agenti negli istanti τ in cui è $q_1(\tau)=0$ provocando una discontinuità su $\dot{q}_1$ (*). Ammette inoltre che il dispositivo sia tale che le azioni elastiche su Ω_2 siano diverse a seconda che è $q_1 < 0$ o $q_1 > 0$. In definitiva Colombo perviene alle equazioni

$$\ddot{q}_1 + 2\,\gamma_1\,\dot{q}_1 + \sigma^2 q_1 = 0; \tag{6.1}$$

$$\ddot{q}_2 + 2\,\gamma_2\,\dot{q}_2 + \alpha_1^2\,q_2 = E \text{ sen } \omega t, \qquad \text{per } q_1 < 0; \tag{6.2}$$

$$\ddot{q}_2 + 2\,\gamma_2\,\dot{q}_2 + \alpha_2^2\,q_2 = E \text{ sen } \omega t, \qquad \text{per } q_1 > 0; \tag{6.3}$$

con γ_1, γ_2, α_1^2, α_2^2 positive. Suppone inoltre che $q_2(t)$, $\dot{q}_2(t)$, $q_1(t)$ siano continue; mentre $\dot{q}_1(\tau)$ è discontinua in tutti gli istanti τ nei quali è $q_1(\tau)=0$.

La discontinuità di $\dot{q}_1$ all'istante τ è però tale che l'energia cinetica di Ω_1 aumenta se q_1 raggiunge l'origine con velocità positiva, diminuisce nel caso contrario. In altre parole, nel primo caso Ω_1 riceve energia da Ω_2; ne cede a Ω_2 nel caso contrario. Si capisce allora che se l'energia ricevuta è superiore a quella ceduta, le perdite energetiche per viscosità possono essere compensate, sicché Ω_1 può compiere oscillazioni persistenti. In altre parole, l'energia prodotta su Ω_2 della forza esterna viene in parte ceduta da Ω_2 ad Ω_1 compensandone così le perdite per viscosità: Ω_1 può così oscillare con il suo periodo proprio, del tutto indipendente dal periodo della forza esterna; appunto perciò l'azione di quest'ultima è detta «azione asincrona».

Cerchiamo ora di descrivere, a grandi linee, il procedimento di Colombo per trattare le equazioni (6.1), (6.2), (6.3) con le citate discontinuità. È bene notare che le discontinuità dipendono da $\dot{q}_2$ e la q_2 varia in modo non facilmente collegato a $q_1(t)$. In seguito, per semplicità, scriveremo q invece di q_1.

Consideriamo l'istante $\tau+0$, cioè un intorno destro dell'istante di di-

---

(*) Il dispositivo di accoppiamento proposto da G. Colombo (e schematizzato in fig. 5) è costituito della leva rigida AOB, fulcrata in O (e suppossta di massa trascurabile), dal filo elastico BCD (scorrevole senza attrito fra i perni C); e da un glifo g lungo il quale può scorrere senza attrito l'estremo A dell'asta AOB. Il moto oscillatorio di Ω_1 (soggetto a forze elastica e viscosa) determina, in corrispondenza dei valori prossimi allo zero della coordinata q_1, il passaggio «brusco» del sistema: asta AOB-filo BCD, dalla posizione I alla posizione II e viceversa, provocando così azioni «quasi» impulsive di Ω_2 su Ω_1 (il carattere «quasi» impulsivo di dette azioni è dovuto al fatto che il tratto inclinato del glifo è quasi ortogonale ai tratti paralleli del glifo stesso; cioè le distanze indicate in figura con q_1^0 sono entrambe prossime allo zero).

scontinuità; e sia $\dot{q}\ (\tau - 0) > 0$ allora sarà $y = \dot{q}\ (\tau + 0) > \dot{q}(r - 0)$ (in seguito scriveremo per brevità τ invece di $(\tau + 0)$. Il moto di Ω_1 nell'intervallo $\tau, \tau + T$ $\left(T = \dfrac{\pi}{\nu}\ ,\ \nu = \sqrt{\sigma^2 - \gamma_1^2}\right)$ sarà per le $(6)_1$ un moto smorzato con condizioni iniziali $q\ (\tau) = 0$, $\dot{q}\ (\tau) = y$ e con semiperiodo T. Si avrà così

$$q\ (t) = \frac{y}{\nu}\ e^{-\gamma(t-\tau)} \text{ sen } \nu\ (t - \tau) \tag{6.4}$$

$$\dot{q}\ (t) = \frac{y}{\nu}\ e^{-\gamma(t-\tau)}\ [\nu \cos \nu\ (t - \tau) - \gamma \text{ sen } \nu\ (t - \tau). \tag{6.5}$$

Siccome $q\ (t)$ è continua, si ha $q\ (t + \tau + T) = 0$, $\dot{q}\,(t + T) < 0$, quindi per effetto della discontinuità dopo l'istante $\tau + T$ l'energia cinetica è diminuita. Nell'intervallo $(\tau + T,\ \tau + 2\ T)$ sarà ancora valida la (6.5), purché in luogo di y si ponga un valore negativo y', $|y'| < y$. All'istante $\tau + 2T$ si ha una nuova discontinuità con aumento di energia cinetica, sicché nell'intervallo $(\tau + 2T,\ \tau + 3T)$ valgono ancora le (6.4), (6.5) con $y > |y'|$. Nel piano delle fasi q, $\dot{q}$ il moto nell'intervallo $(\tau,\ \tau + 2T)$ è rappresentato dalla curva della figura 6. Il tratto ABC rappresenta il moto

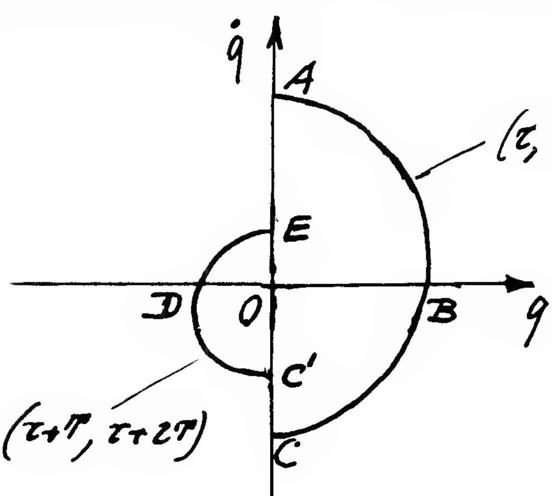


Figura 6

di Ω_1 nell'intervallo $(\tau,\ \tau + T)$ con $OC < OA$ per effetto dello smorzamento; CC' il salto di $\dot{q}$ all'istante $\tau + T$; $C'DE$ rappresenta il moto di Ω_1 dell'intervallo $(\tau + T,\ \tau + 2T)$; EE' il salto di $\dot{q}$ nell'istante $\tau + 2T$. Dopo $\tau + 2T$, il moto si ripete con le stesse caratteristiche; cioè il moto

stesso è una successione di moti smorzati continui per $q\ (t)$, discontinui per $\dot{q}\ (t)$ negli istanti $\tau + nT$ (n intero). Come si è detto, l'effetto della discontinuità è diffieile da apprezzare: esso dipende infatti anche da $q_2(t)$. Comunque, Colombo prova non solo che $q\ (t)$ si annulla negli istanti $\tau + 2nT$, $\tau + (2n+1)\ T$ (n intero) e che negli intervalli $(\tau + 2\ nT$, $\tau + (2n+1)\ T)$ la $q\ (t)$ è positiva e con un solo massimo, mentre nell'intervallo successivo è negativa con un solo minimo: proprietà, questa, del tutto ovvia, in quanto proprietà delle oscillazioni smorzate; ma Egli riesce anche a provare che quei massimi e minimi sono sempre compresi tra due numeri fissi $y_1 y_2$ indipendenti dalle soluzioni ora considerate, cioè $y_1 < q\ (t) < y_2$. Il moto ha dunque carattere oscillatorio il cui periodo può ritenersi, come nel moto smorzato, uguale a $2T$ cioè indipendente dal periodo $\dfrac{2\pi}{\omega}$ della forza agente su Ω_2. Però l'ampiezza del moto è variabile, ma limitata; sicché il moto presenta carattere di stabilità. Colombo chiama questo moto: «moto oscillatorio isocrono».

È bene notare che dopo gli istanti $t_0 = \tau + (2n+1)\ T$ l'energia cinetica diminuisce; potrebbe quindi succedere che $\dot{q}^2$ risultasse negativo, il che ovviamente è assurdo: allora, dopo t_0, $q\ (t)$ rimane sempre uguale a zero; cioè Ω_1, dopo t_0 rimane in quiete nella posizione di equilibrio.

7. Colombo (1958) ha anche sviluppato una teoria del regolatore di Bouasse e Sarda.

Anche questo dispositivo è a due gradi di libertà, perché è determinato dall'angolo θ di rotazione dell'albero a partire dalla posizione in cui il gomito è orizzontale e dalla coordinata x del baricentro G della massa M (cfr. fig. 7).

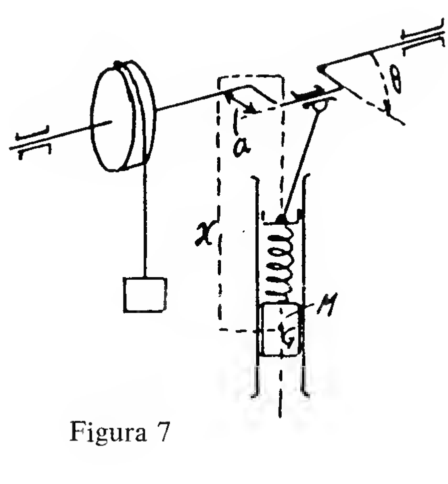


Figura 7

Le equazioni del moto sono state stabilite dal Rocard (Dynamiqye générale des vibrations, Masson, 1949, pp. 312-315); equazioni che però Colombo completa, aggiugendo un termine dissipativo nelle equazioni del moto dell'albero; però Egli semplifica tale equazione, supponendo molto grande il momento d'inerzia della massa rotante. Giunge così alle equazioni:

$$\ddot{\theta} = \epsilon\ [\alpha + k\ (x - a \text{ sen } \theta)\ a \cos \theta - D\dot{\theta}] \tag{7.1}$$

$$\ddot{x} + 2\ \lambda\omega_0\dot{x} + \omega_0^2 x = \epsilon\ (\omega_0^2\ a \text{ sen } \theta + g) \tag{7.2}$$

dove ϵ è il solito parametro positivo che si suppone molto piccolo; α, k, λ, D sono costanti positive; $D\dot{\theta}$ è il termine dissipativo introdotto da Colombo, a il raggio di manovella del gomito; infine ω_0 è la pulsazione che il moto armonico di M avrebbe se la molla fosse distaccata dall'albero, cioè con l'estremo B di essa supposto fisso (oscillazioni libere).

Colombo risolve le (7.1) e (7.2) ponendo

$$x = x_0 + \epsilon\ x_1, \quad \vartheta = \omega t + \epsilon\ \theta_1 \tag{7.3}$$

(dove x_0 e ω sono costanti mentre x_1 e θ_1 sono funzioni del tempo); trascurando termini $O\ (\epsilon^2)$ perviene a equazioni facilmente risolvibili.

Esiste un regime in cui l'albero ruota, salvo piccoli scarti, con velocità angolare costante ω^*. Posto $z = \dfrac{\omega}{\omega_0}$. Egli dimostra che $z^* = \dfrac{\omega^*}{\omega_0}$ è definita dai punti di intersezione delle due linee

$$y_1(z) = \alpha - D_2 z \tag{7.4}$$

$$y_2(z) = \frac{\lambda\ k\ a^2 z}{(1 - z^2)^2 + 4\ \lambda^2 z^2} \tag{7.5}$$

dove α è una costante proporzionale a m, massa del corpo che discende e che fa ruotare l'albero; e che il moto può ritenersi stabile se è:

$$y_1'\ (z^*) < y_2'\ (z^*). \tag{7.6}$$

La $y_1(z)$ è una retta di coefficiente angolare $-D$ che al crescere di α si allontana dall'origine rimanendo parallela alla propria direzione ini-

ziale. La $y_2(z)$ è una curva simmetrica rispetto all'origine, rappresentata nel primo quadrante della fig. 8.

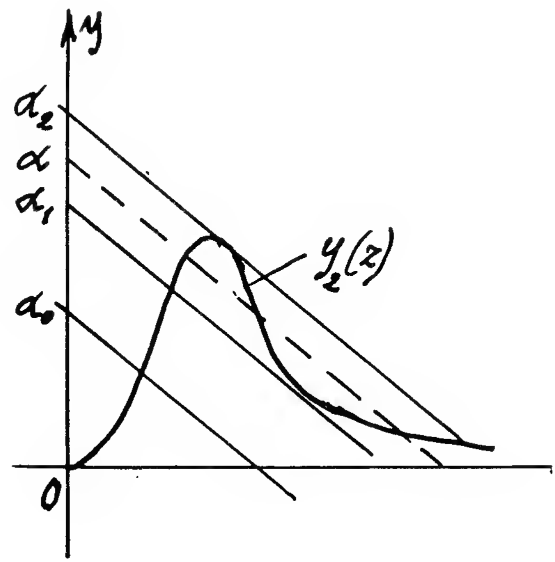


Figura 8

Ora, fino a che α resta minore di α_1, (α_1 è il valore di α per cui $y_1(z)$ diventa per la prima volta tangente a $y_2(z)$, vi è un solo punto z^* intersezione di $y_1(z)$ e $y_2(z)$; in questo punto è:

$$y_1'(z^*) = -D, \; y_2'(z^*) > 0$$

(la $y_2(z^*)$ è nel ramo crescente della curva $y = y_2(z)$); quindi la (7.6) é soddisfatta e $z^* = \dfrac{\omega^*}{\omega_0}$ caratterizza la pulsazione ω^* di un moto di rotazione stabile.

Sia poi α_2 l'altro valore di α per cui la curva $y_1(z)$ è di nuovo tangente a $y_2(z)$. Allora, la retta $y_1(z)$ con $\alpha_1 < \alpha < \alpha_2$ incontra la $y_2(z)$ in tre punti $z_1^* < z_2^* < z_3^*$. Ora, fino a che z_1^* resta nel tratto ascendente della $y_2(z)$ la (7.6) è ancora soddisfatta e tale resta per continuità fino a che α raggiunge il valore α_2 dove è $y_1'(z_1^*) = y_2'(z_1^*)$. Invece in z_2^* l'angolo che la tangente in forma con l'asse z, è ottuso e minore in ampiezza dell'angolo che la retta $y_1(z)$ forma con l'asse z: sussiste quindi la disequazione:

$$y_2'(z_2^*) < y_1'(z_2^*)$$

che contraddice la (7.6). Il punto z_2^* è perciò instabile. Nel punto z_3^* invece, vale ancora la (7.6) e il moto è pertanto stabile. Se è poi $\alpha > \alpha_2$ si ha una sola soluzione sempre stabile.

Ragionando come nel paragrafo 5. potrebbe manifestarsi in questo caso, al variare di α, ossia al variare di m, ancora un fenomeno di jump e di isteresi; ma su ciò qui non insisteremo.

8. Come è stato già accennato, Colombo ha studiato anche questioni riguardanti sistemi a un solo grado di libertà; di ciò parleremo in questo paragrafo e nei seguenti. Cominceremo con alcune ricerche sulle oscillazioni non lineari di combinazione (1957/1), (1957/2), riportando un elegante esempio meccanico citato da Colombo (1957/1). Consideriamo un corpo di massa M collegato con una molla di costante k e con una resistenza non lineare al supporto C_1 e con una resistenza passiva collegata al supporto C_2 (fig. 9). Rispetto ad un sistema inerziale C, il

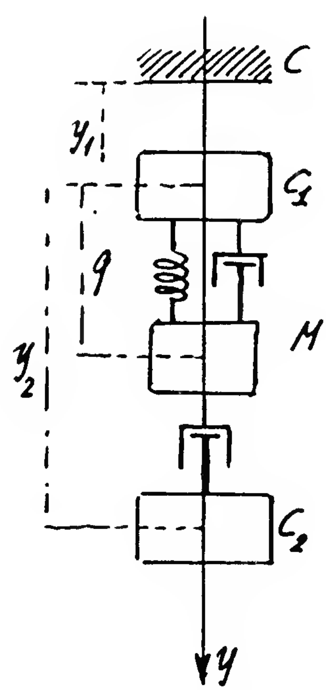


Figura 9

moto di M e dei supporti sia di traslazione parallela ad una retta y, sicché sia y_1 la coordinata di C_1, y_2 quella di C_2 rispetto a C_1, q la coordinata di M rispetto a C_1 (sicché rispetto a C la coordinata di M vale $q+y_1$). Supporremo inoltre:

$$(8.1) \qquad y_1 = e \text{ sen } \omega t, \; y_2 = h - b \cos(\omega t + \psi).$$

Per quanto si è detto, le forze agenti su M dovute a C_1 ed a C_2 sono

$$(8.2) \qquad \phi_1 = -kq - (\alpha_1 \dot{q} + \beta_1 \dot{q}^3), \; \phi_2 = -\alpha_2(\dot{q} - \dot{y}_2) - \beta_2(\dot{y} - \dot{y}_2)^3$$

l'equazione del moto di M è pertanto

$$(8.3) \qquad M\ddot{q} + kq + \alpha_1 \dot{q} + \beta_1 \dot{q}^3 + \alpha_2[\dot{q} - b\omega \text{ sen } (\omega t + \psi)] +$$
$$- \beta_2[\dot{q} - b\omega \text{ sen } (\omega t + \psi)]^3 = M\, e\, \omega^2 \text{ sen } \omega t$$

(va osservato qui che se tutte le grandezze α, β fossero nulle; si avrebbe un moto armonico di pulsazione $\omega_1 = \dfrac{k}{M}$ quindi il sistema non lineare si comporterebbe come un sistema nel quale si sovrappongono due frequenze diverse). Dividendo per M, ponendo $E = \omega^2 e$, $\epsilon = \dfrac{1}{M}$ e scegliendo in modo opportuno l'unità di tempo, si perviene all'equazione:

$$(8.4) \qquad \ddot{q} + q - E \text{ sen } \omega t = \epsilon\, f\,[q, \dot{q}, \text{ sen } \omega t, \text{ sen } (\omega t + \psi)]$$

con ovvio significato dei simboli al secondo membro e nel quale ϵ é un piccolo parametro positivo.

Per risolvere la (8.4) usiamo qui, a differenza di Colombo, il metodo K.B., ritrovando però i di Lui risultati. Poniamo pertanto:

$$(8.5) \qquad q = a \text{ sen } (t + \theta) + \frac{E}{1-\omega^2} \text{ sen } \omega t \; ,$$

$$\dot{q} = a \cos (t + \theta) - \frac{E\omega}{1-\omega^2} \cos \omega t$$

dove a e θ sono funzioni di t. Con questa posizione, e considerando a e θ costanti, il secondo membro di (8.4) diventa funzione di t e di $\tau=\omega t$; e quindi potremo scriverlo nella forma $F(t, \tau)$ essendo F una funzione periodica di periodo 2π sia rispetto a t che rispetto a τ. Allora, con i soliti procedimenti a suo tempo richiamati, si trova:

$$\frac{da}{dt}=\epsilon F(t, \tau) \cos (t+\theta); \tag{8.6}$$

$$\frac{d\theta}{dt}=-\epsilon F(t, \tau) \operatorname{sen} (t+\theta). \tag{8.7}$$

Sviluppiamo ora, sempre mantenendo a e θ costanti, $F(t, \tau)$ in serie doppia di Fourier; serie che ha però qui soltanto un numero finito di termini diversi da zero; e ciò in quanto $F(t, \tau)$ è un polinomio trigonometrico nelle funzioni sen $(t+\theta)$, cos $(t+\theta)$, sen ωt, sen $(\omega t+\psi)$.

Se ω è molto diversa da 1 cioè se è $|\omega-1| >> \epsilon$; o, più in generale ogni termine della serie che ha per argomento $(n+m\omega)\, t$ (con m e n numeri interi positivi o negativi) è $|n+m\omega| >> \epsilon$ (si evitano in tal modo i piccoli denominatori), si può affermare che, con errore trascurabile, si possono sostituire ai secondi membri di (8.6) e (8.7) i rispettivi valori medi scrivendo cioè:

$$\frac{da}{dt}=R(a)=\frac{\epsilon}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi} F(t, \tau) \cos (t+\theta)\, dt\, d\tau, \tag{8.8}$$

$$\frac{d\theta}{dt}=\S\, \frac{\epsilon}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi} F(t, \tau) \operatorname{sen} (t+\theta)\, dt\, d\tau. \tag{8.9}$$

Come esempio consideriamo il caso in cui sia $b=0$, $\alpha_2=0$, $\beta_2=0$; e sia inoltre $\alpha_1\dot{q}-\beta_1\dot{q}^3$ sostituito da $(\alpha-\beta q^2)\dot{q}$: la (8.4) assume la forma:

$$\ddot{q}+q-E \operatorname{sen} \omega t=\epsilon\, (\alpha-\beta\, q^2)\, \dot{q} \tag{8.10}$$

interessante anche perché rappresenta un circuito di Van der Pol nel quale

sia inserita una forza elettromotrice alternativa di pulsazione ω.

Si ha quindi, ponendo ancora $\tau=\omega t$:

$$F(t,\ \tau)=\epsilon\left\{\alpha-\beta\left[a\ \text{sen}\ (t+\theta)-\frac{E\,\text{sen}\,\omega\,t}{1-\omega^2}\right]^2\right\} \tag{8.11}$$
$$\left[a\ \cos\ (t+\theta)-\frac{E\ \omega\cos\omega\,t}{1-\omega^2}\right].$$

Sostituendo in (8.8) e (8.9) ed eseguendo gli integrali si trova:

$$\frac{da}{dt}=R\ (a)=\epsilon\alpha\ \frac{a}{2}\left(1-\frac{\beta}{\alpha}\ \frac{a^2}{4}-\frac{\beta}{\alpha}\ \frac{E^2}{2\ (1-\omega^2)^2}\right); \tag{8.12}$$

$$\frac{d\theta}{dt}=0\Rightarrow\theta=\theta_0.$$

In condizione di regime deve essere ovviamente $\dfrac{da}{\mathrm{dt}}=0$; cioè ricordando che è $\dfrac{\beta}{\alpha}=\dfrac{4}{a_0^2}$ deve aversi:

$$\frac{\alpha\ a}{2}\left(1-\frac{a^2}{a_0^2}-\frac{4}{2\ a_0^2}\ \frac{E^2}{(1-\omega^2)^2}\right)=0. \tag{8.13}$$

Si ha quindi regime se è $a=0$ oppure se è:

$$a^{*2}=\left[a_0^2-\frac{2\ E^2}{(1-\omega^2)^2}\right]. \tag{8.14}$$

Studieremo ora la stabilità della soluzione di (8.10) espressa da (8.5) nella quale in luogo di a si pongano i valori or ora trovati; basterà dimostrare che se a devìa di poco dalla condizione di regime, tende a ritornarvi. Infatti, sviluppando $R\ (a)$ in un intorno del primo ordine di a^*, si ha

$$R(a) = R(a^*) + \frac{dR(a^*)}{da}(a - a^*)$$

$$\frac{dR(a^*)}{da} = -\alpha\epsilon\frac{a^{*2}}{a_0^2} < 0.$$

Ricordando quindi la (8.12), risulta che se è $a - a^* > 0$ è anche $\frac{da}{dt} < 0$ (cioè decrescente); se è invece $a - a^* < 0$ si ha $\frac{da}{dt} > 0$ (cioè a è crescente). In ogni caso, dunque, se a viene spostata dal valore di regime, tende a ritornarvi.

Più in generale, si può dire che condizione sufficiente per la stabilità è che sia $\frac{dR(a^*)}{da} < 0$: risultato questo ottenuto da Colombo per altra via.

9. Riprendiamo ora l'equazione di Van der Pol delle oscillazioni forzate e supponiamo che sia $\alpha = \beta$: caso questo cui ci si può sempre ricondurre con un opportuno cambiamento di incognita.

Scriveremo dunque:

$$\ddot{q} - \alpha(1 - q^2)\dot{q} + q = f(t). \tag{9.1}$$

Ci proponiamo ora di vedere per quali valori di $f(t)$ la (9.1) può essere soddisfatta dalla funzione $q = 2 \operatorname{sen} t$; si ha subito con semplice calcolo:

$$\begin{aligned} f(t) &= -\alpha(2\cos t - 8\operatorname{sen}^2 t\cos t) = \\ &= -2\alpha(\cos t - 2\operatorname{sen} t\operatorname{sen} 2t) = \\ &= -2\alpha(\cos t - \cos t + \cos 3t) = -2\alpha\cos 3t. \end{aligned} \tag{9.2}$$

Si può pertanto affermare che l'equazione (9.1) con $f(t)$ uguale a $-2\alpha\cos 3t$ ammette la soluzione esatta $2 \operatorname{sen} t$, cioè sottoarmonica di ordine tre.

Colombo (1953) ha però dimostrato che la citata soluzione è, per α abbastanza piccola, instabile.

Siccome col metodo sopra esposto è facile costruire soluzioni sottoarmoniche di equazioni non lineari, è da attendersi che tali soluzioni non siano stabili e non abbiano pertanto alcun interesse pratico.

Colombo osserva inoltre che l'equazione alle variazioni della (9.1) coincide con l'equazione alle variazioni della stessa (9.1) quando in essa sia posto $f(t)=0$. In questo caso la funzione 2 sen t risulta soluzione approssimata stabile. Sorge perciò il dubbio che la stabilità della soluzione approssimata non assicuri la stabilità della soluzione esatta corrispondente. La questione è di indubbio interesse; ma non mi risulta che sia stata fin'ora risolta. D'altra parte Colombo, nei propri lavori, parla sempre soltanto di «stabilità approssimata».

10. A proposito di oscillazioni sottoarmoniche è di rilevante interesse anche la nota $(1954)_1$. In essa Colombo considera un circuito costituito da una bobina a nucleo di ferro in serie ad una resistenza R e ad una capacità C alimentato da una forza elettromotrice periodica $E(t)$ di periodo $2T$ (fig. 10). Sia $q(t)$ la carica sull'armatura positiva del conden-

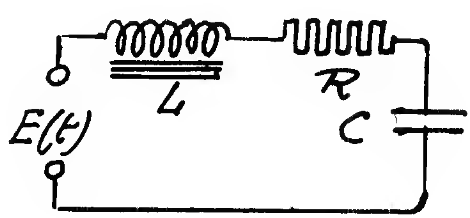


Figura 10

satore; $i=\dot{q}(t)$ l'intensità della corrente e $\phi(\dot{q})$ il flusso del vettore induzione nella bobina. Per la presenza del ferro, anche trascurandone l'isteresi e soprattutto per effetto della saturazione, il flusso ϕ non è funzione lineare di $\dot{q}$.

L'equazione a cui soddisfa la q è allora la seguente:

$$\frac{d\phi}{d\dot{q}}\ddot{q}+R\dot{q}+\frac{1}{C}q=E(t)\,. \tag{10.1}$$

Colombo schematizza la funzione $\phi(\dot{q})$ ammettendola simmetrica rispetto all'origine; e, nel primo quadrante, tale che quando è $i \leqslant I_s$ (I_s

corrente di saturazione), la ϕ è proporzionale a $\dot{q}$, cioè si ha $\phi\ (\dot{q}) = L_1\dot{q}$; per $i > I_s$ la funzione $\phi\ (\dot{q})$ è ancora lineare ma è della forma $\phi\ (\dot{q}) = L_2\dot{q} + c$, dove c è una costante (fig. 11) mentre le induttanze L_1

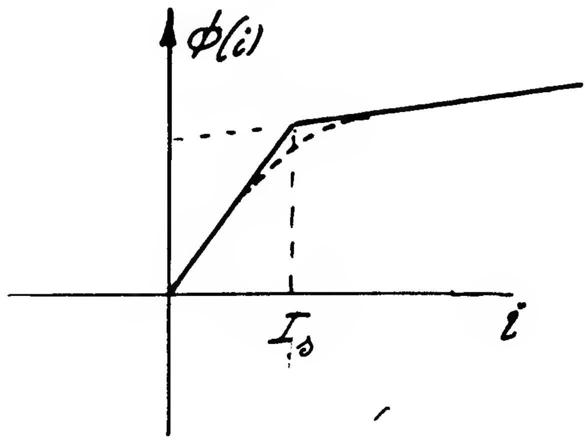


Figura 11

e L_2 sono positive. Ammette inoltre che la $E\ (t)$ sia non solo periodica ma verifichi le condizioni:

$$E\ (t) = E^*\ (E^* \text{ costante}) \text{ per } \forall\ t \in (0,\ T)$$

$$E\ (t) = -E^* \text{ per } \forall\ t \in (T,\ 2\ T).$$

La (10.1) diventa allora

$$(10.2)_1 \qquad L_1\ddot{q} + R\dot{q} + \frac{1}{C}\ q = E\ per\ \forall\ |\dot{q}| < I_s$$

$$(10.2.)_2 \qquad L_2\ddot{q} + R\dot{q} + \frac{1}{C}\ q = E \text{ per } \forall\ |\dot{q}| > I_s.$$

Le $(10.2)_1$, $(10.2)_2$ sono equazioni lineari, però tali che quando è $\dot{q} = I_s$, e q, $\dot{q}$ devono risultare continue. La trattazione del problema presenta difficoltà del tipo di quelle incontrate per le (6.1) e (6.2); Colom-

bo riesce però, in un primo tempo, trascurando R, a mostrare che le $(10.2)_1$ e $(10.2)_2$ ammettono una soluzione periodica di periodo 6 T; poi, con considerazioni topologiche, riesce a dimostrare l'esistenza di qualche soluzione anche supponendo che la $\phi(\dot{q})$ non abbia in $\dot{q}=I_s$ un punto angoloso; cioè che si abbia in corrispondenza di I_s un raccordo tra i due tratti lineari della $\phi(\dot{q})$. Colombo suppone inoltre che la $E(t)$ sia continua, in quanto egli ammette che in un piccolo intervallo intorno agli istanti $2nT$, $(2n+1)T$ la $E(t)$ passi con continuità da $+E^*$ a $-E^*$, e viceversa. Nel circuito si può quindi avere una corrente sottoarmonica di ordine tre e questo risultato è conforme all'esperienza.

11. Colombo (1955/3) si è occupato anche di sistemi a ritardo. Nella nota or ora citata Egli considera l'equazione

$$\ddot{x}+\epsilon Dx+x_\epsilon-\beta x_\epsilon^3=0 \qquad x_\epsilon=x(t-\epsilon) \tag{11.1}$$

nella quale ϵ è il solito parametro molto piccolo, in virtù del quale può scriversi:

$$x_\epsilon=x(t)-\epsilon\dot{x}(t)+0(\epsilon^2). \tag{11.2}$$

La (11.1) diventa pertanto

$$\ddot{x}+\epsilon(D-1)\dot{x}+x-3\beta\epsilon x\dot{x}-\beta x^3+O(\epsilon^2)=0 \tag{11.3}$$

Colombo riesce a dimostrare che sotto opportune condizioni e se è $D<1$, la (11.3) ammette soluzioni oscillatorie prossime ad alcune soluzioni periodiche dell'equazione

$$\ddot{x}+x-\beta x^3=0. \tag{11.4}$$

12. Più di recente, Colombo (1959) ha dimostrato il seguente teorema: data l'equazione

$$\alpha(x)\ddot{x}+\frac{1}{2}\alpha'(x)\dot{x}^2+g(x)=\epsilon f(x,\dot{x},\epsilon) \tag{12.1}$$

che, per $\epsilon=0$ ammette l'integrale primo

$$(12.2) \qquad \alpha(x)\frac{\dot{x}^2}{2} + G(x) = G(x_0) \text{ (con } G(x) = \int g(x)\,dx\text{)},$$

suppone che per ogni x_0 compreso nell'intervallo (x_1, x_2) le soluzioni di (12.2) siano periodiche in modo che nel piano delle fasi $\pi(x_1, x_2)$ rappresentino un insieme di cicli limite stabili γ_{x_0}. Costruisce poi l'integrale, funzione della sola x_0

$$(12.3) \qquad J(x_0) = \int_{\gamma_{x_1}} f(x, \dot{x}, \epsilon)\,dx$$

e dimostra che se esiste un x_0^* tale che sia

$$(12.4) \qquad J(x_0^*) = 0, \quad \frac{dJ(x^*)}{d\,x_0} < 0,$$

allora, per ϵ sufficientemente piccolo, esiste una soluzione della (12.1) rappresentata nel piano delle fasi da un ciclo limite stabile. Questo teorema potrebbe forse applicarsi utilmente per determinare le soluzioni periodiche stabili della (11.3).

Ringrazio vivamente il prof. Luigi Caprioli per il Suo validissimo aiuto.

BIBLIOGRAFIA
(Lavori di Giuseppe Colombo citati in questa Memoria)

(1950/1) *Sull'equazione differenziale non lineare del terzo ordine di un circuito oscillatorio a triodi.* Ren. Sem. Mat. Univ. Padova (XIX pp. 134-140).

(1950/2) *Sulle oscillazioni non lineari in due gradi di libertà.* Ren. Sem. Mat. Univ. Padova (XIX pp. 413-441).

(1952/1) *Sopra un sistema non lineare in due gradi di libertà.* Idem (XXII pp. 63-98).

(1952/2) *Sopra un fenomeno di isteresi oscillatoria*. Idem (XXI pp. 370-382).

(1953/1) *Sopra un singolare caso che si presenta in un problema di stabilità in Meccanica non-lineare*. Idem (XXII pp. 123-133).

(1954/1) *Sulle oscillazioni forzate di un circuito comprendente una bobina a nucleo di ferro*. Idem (XXIII pp. 407-421).

(1954/2) *Sui sistemi autonomi di ordine superiore al secondo*. Conferenza tenuta al I° Corso CIME, 1954.

(1955/1) *Sopra il fenomeno dell'azione asincrona*. Ren. Sem. Mat. Univ. Padova (XXIV pp. 353-395).

(1955/2) *Moti di regime di un sistema non lineare autonomo in due gradi di libertà, con debole accoppiamento capacitivo*. Idem (XXIV pp. 400-420).

(1955/3) *Oscillazioni persistenti di un sistema non lineare dissipativo dovute al ritardo della forza di richiamo*. Annali Univ. Ferrara (IV pp. 33-50).

(1957/1) *Sopra un notevole fenomeno nel campo delle vibrazioni non lineari di combinazione*. Rend. Acc. Naz. Lincei (XXII pp. 726-730).

(1957/2) *Sulle oscillazioni non lineari di combinazione*. Rend. Sem. Mat. Univ. Padova (XXVII pp. 162-175).

(1958) *Teoria del regolatore di Bouasse e Sarda*. Idem (XXVIII pp. 338-347).

(1959) *Sulla determinazione analitica delle soluzioni periodiche dei sistemi non lineari autonomi*. Rend. Acc. Naz. Lincei (XXVI pp. 662-664).

Safety in astronautic activities

Ettore ANTONA *

1. FOREWORD

This paper will not be a discussion on safety levels in so far obtained but an attempt to contribute to the definition of logical and methodological basis for further scientific improvements of safety concepts.

The definition of «safety», [1], is: «The conservation of human life and its effectiveness, and the prevention of damage to items, consistent with mission requirements».

Safety as defined in [1] is an engineering problem, without any specification in a deterministic or probabilistic problem.

Safety has had probabilistic aspects since the early of aeronautics, particularly as allowable material characteristics and flight loading definitions are concerned.

During the evolution of aeronautical and space technology, the importance of probabilistic aspects grew strongly up to be involved also in the (statistical) analysis of the probability of failure of the various components (engines included) and, more recently, in the analysis of structure failure loads, crack propagation, Non Destructive Inspection method capabilities, and so on.

The overall safety level results from a very large number of aspects (see for instance fig. 1,1) many of which from a theoretical point of view can be analysed only by means of statistical concepts. Therefore, we can say that safety is the probability of having not an hazard (i.e. an event that potentially can give damages) during a given employment time (or number of missions) with given operative conditions. As some technics are concerned the overall hazard probability to be obtained is so remote that it can be practically disregarded (safety practically equal to the unity). This notwithstanding, from a general point of view a determination of the obtained safety level could be made only in a statistical sense.

On the other hand, in practical design operations, safety is pursued in each particular field by means of precautions that following each a particular considered problem result of a probabilistic nature, or of a deterministic nature even if deterministic criteria are very often ap-

plied to values or data of a probabilistic origin. Thus the obtained overall safety is very difficult to be calculated, particularly when the hazard deriving from human errors, weather and meteorological events must be included in the evaluation. Furtherly, the more safety is near to the unity,the more its evaluation is difficult. In the opinion of A., among the aims of the future theoretical and experimental studies, besides the improvement of particular and overall safety levels, there shall be the adeguacy of theoretical formulations and practical data to the purpose of evaluating safety as an overall figure.

Hazards	Reentry impossibility	Reentry not safe	Ambiental conditions	Not controllable trim.	Launch system failure
Hazards causes	— Engine failure	— Structural failure	— Toxicity	— Engine failure	— Fire
	— Structural failure	— Control loss	— High radiation level	— Control loss	— Structural failure
	— Control loss	— Protection system failure	— High temperat. level		— Engine failure
	— Protection system failure		— Explosions		— Protection system failure

Fig. 1,1 - Crew safety during space missions. (A sample analysis of hazard sources).

The existence of an overall safety level, resulting from all safety precautions considered as a whole, even if not early evaluable at the present status of our knowledge, and the acceptance of such a level as the target of an entire set of prescriptions can offer a clear way of thinking at the present moment of the space activity.

Safety problems are present during each one of the three phases: design, construction and use of a space system. Within the design phase the two following phases must be taken into account. From a general point of view safety appears an amount of conditions to be imposed to the design. In fact a design can be conceived as a (mathematical) problem of optimum — of an objective characteristic — under several conditions

a certain number of which are connected with the required safety level (see fig. 1,2).

x_i = design un knowns	$(i=1, 2, \ldots)$
y_j = conditional parameters	$(j=1, 2, \ldots)$
$Q_h = Q_h(x_i)$ = Characteristics to be minimized	
DESIGN	$\begin{cases} Q(x_i)\varphi_p(Q_h) \\ y_j \geqslant \alpha_j \ [f_q(y_j) \geqslant \alpha_q] \end{cases}$

Fig. 1,2 - Design as conditioned optimum problem.

Taking into account the above considerations, this lecture is mainly concerned with the uphold of the following three thesis:

1) Safety in aerospace (activities) is such a body of knowledges and involves such an amount of theoretical basis that by now it forms a fundamental research and teaching matter like the traditionally involved ones.
2) Safety, in order to be achieved, requires both well defined criteria, based on the knowledge of the phisical behaviour, and methodologies to control and assure their application.
   The good safety level already obtained in the aerospace activities must urge us to advance in the study both of proper criteria and suitable methodologies.
3) The safety level to be achieved must be determined by authorities in a way that may ensure that the wishes of the community are taken into account.
   Scientist and engineer roles are to put safety problems into a correct form by means of clear and comprehensible statements and to achieve such prescribed safety level and to demonstrate their achievement in the design.

Several arguments can be considered that sustain the necessity of safety levels determined in a way that can ensure the approval of the community.

First of all space activities concern life and integrity of several human beings, both directly connected with the program or not.

Secondly their success has direct consequences on the credibility of the country or of the country group that have realized them.

Furtherly, space activities require great quantities of human and finantial resources and common people often wonder if these resources are correctly employed in such kind of activities.

2. GENERAL CONCEPTS AND STRATEGIES FOR SAFETY

2.1 Definitions and basic guidelines

Safety, as previously said in a brief criticism of its most general definition (see para. 1)), can be referred to as the probability of having not an hazard (i.e. an event that potentially can give damages) during a given employment time (or number of missions) with given operative conditions.

A sample of classification of the hazard severity categories is presented in fig. 2,1, where catastrophic, critical and marginal hazards are considered. A general way of thinking in order to obtain safety, applicable to which ever design problem, is summarized in fig. 2,2, where 1) elimination, 2) minimization and 3) control of the hazard are considered as actions to which one can apply in such an order of precedence. The practical application of such concepts implicates the use of appropriates devices, whose probability of success are to be investigated and taken into account in safety analyses. This is particularly important in the case of a human error possibility, where a modern way of thinking is its avoidance by means of automatism or monitors.

I) Catastrophic hazards.
May cause loss of life or loss of important parts of the system.

II) Critical hazards.
May cause severe illness to people or damages to important parts of the system.

III) Marginal hazards.
May cause minor illness or damages.

Fig. 2,1 - Hazard severity categories (sample of classification).

1) Hazard elimination by removed of hazard sources and hazardous operations by applications of appropriate design measures.

2) Minimization of hazards through the selection and incorporation of appropriate design measures.

3) Control of hazards:

   a) by the use of appropriate safety devices:
   b) through the use of warning devices;
   c) through the use of emergency devices;
   d) through adherence to a planned maintenance and repair schedule.

Fig. 2,2 - Order of precedence in order to assure that safety is designed into the system.

Safety prescriptions give useful criteria both for the minimization and for the control of the hazard, that, with the success probability of the various components, contribute to the safety level. Samples of qualitative and of quantitative prescriptions are reported in fig. 2,3 and fig. 2,4, derived from STS payload and from «Columbus» prescriptions respectively. It is easy to understand the influence that prescriptions of such a nature can have on the overall safety.

(From prescriptions for STS payloads)

Control of hazardous functions

a) A function that could result in a catastrophic hazard must be controlled by a minimum of three independent inhibits (other prescriptions follow).

b) A function that could result in a critical hazard must be controlled by a minimum of two independent inhibits (other prescription follow).

Fig. 2,3 - Samples of safety prescriptions of qualitative nature.

(From COLUMBUS prescriptions)

Resistence to micrometeoroids/debris

a) *Pressurized modules*: the probability based on one year as regards to penetrations able to cause leakages resulting in a drop from 1013 mBAR to 700 mBAR within less than two minutes must be smaller than 0.0005 (*).

b) *Pressure vessels*: the probability based on one year penetrations must be smaller than 0.001.

c) *Other structures*: the probability based on one year, of having significant functional/operational degradations, must be smaller than 0.01.

(*) Repair capability/procedures/equipment shall be provided with to ensure recovery from a failed condition.

Fig. 2,4 - Samples of safety prescriptions of quantitative nature. (From COLUMBUS prescriptions).

2.2 Overall safety

As each other system subjected to safety requirement, a «space system» can be considered a set of components (let m indicate their number), each having a proper success probability p_i in a given time or in a defined mission, $1-p_i$ being, obviously the failure probability. The success of the entire set, as safety is concerned, (i.e. the lack of hazards) is assumed when several subsets of components, where a single element can belong to more subsets, have success.

A clear way of thinking is obtained if the entire field of probability is divided into events that are considered as simple (let n be their number) where from a logical point of view each of them is independent from the others. In other words, each events has a probability given by the product of the probability of its components and the sum of all the probability of the various event is the unity.

If we consider two events — success and failure — for each component, we can introduce «a space of the events» (see [26]) whose number n is given by

$$n=2^m.$$

In other words in each event all components are considered in function or not. The probability of an event is the product of the probability of the status of each component and the sum of the probability of all the events is equal to unity. The $n = 2^m$ events can be divided into two fields. One contains the events that can assure the success and the other the events that give the failure of the entire system. If $m = 1, 2, \dots f$ are the components in function and $m = f + 1, \dots m$ are the failured ones, the probability of the event is:

$$P_e = p_1 p_2 \dots p_f \cdot (1-p)_{f+1} \dots (1-p)_m \qquad (e = 1 \dots n)$$

and

$$\sum_{e=1}^{n} P_e = 1.$$

In such a manner the probability field coincides with the «space of events» and the determination of the success probability of the system is given by the sum of the probability P_{e_f} of the events that assure the success:

$$P_s = \Sigma \, P_{e_f}.$$

2.3 Structural safety

In the application, and particularly in the design application, structural safety is approached where deterministic and probabilistic components play roles of various importance due to the knowledge of the structural behaviour and loading and environmental conditions.

A complete probabilistic approach seems to be still far in the future. At present it seems yet to be demonstrated that such an approach could be the best for a significant judgement of the safety margins of a certain structure in certain employ conditions. In any case, if several theoretical concepts for a complete probabilistic approach are developed or under way of development, the necessary experimental data are far from a suitable availability.

The safety level of a structure is obtained by means of a series of criteria and requirements that must be considered as a whole. In particular it is to be emphasized that the criteria, such as, for instance, the ultimate factor of safety, are not exhaustive in themselves of the safety level, unless they are accompanied by the quality control or product assurance requirements.

The evolution of structural safety concepts from the early statements up today has been promoted mainly by improvements in the knowledge of the phenomena involved in the flight and by the fast enlargement of the environmental and loading conditions, in which the flight structure has been involved. To this enlargement both airplane and missile structures have been interested in a unified process of safety thinking clarification, even though either types of structures have been analyzed so far with different numerical values of the various safety factors.

Beside the usual uncertainty on several quantities involved, such as material strength properties, geometrical dimensions, and so on, the simplest aerospace structural problem, i.e. the static strength of a structure, has a wide uncertainty on the applied loads and, in particular, on their distribution.

Thus, a fully deterministic approach, that would imply very narrow probability distributions of all the quantities involved, is not possible. Since the early aeronautical designs structural safety has been obtained from deterministic safety factors (implicit or explicit), which mainly reduced the probability of having loading conditions more severe than the ones used in the absence of reliable probabilistic description of the loading conditions and of the strength capabilities. Even though strengths in the static case show no great scatter because of the requirements in the material and manufacture acceptance.

When probabilistic descriptions of loads or strengths are available, an approach based also on probabilistic criteria is possible. This becomes necessary when either or both the data have a wide scatter.

A fully probabilistic approach where a total failure probability in a given time is calculated — by taking into account all the environmental and loading conditions to be withstood by the structure — is not applied at present, due to the inadequacy of several necessary statistical data.

Nevertheless, an evolution toward the probabilistic approach is observed in spite of the enlargement of the physical problems involved and requested data, (see for instance [2]).

The more recent of such enlargements is connected with fracture mechanics concept, which had to be included in the aerospace structure analysis. Fracture mechanics studies the residual strength of structures with defects and the growing of cracks or flaws from their initial dimensions up to critical ones at which fracture starts. The processes involved by these two aspects are very much complicated, but a body of knowledge has been established through which designers can approach the problem.

In such an evaluation, a primary role is played also by N.D.I. methods. They give the possibility of measuring cracks, flaws and other

defects, and their capabilities are to be considered in a statistical way.

Theoretical analysis, tests and non-destructive inspection involved in the fracture mechanics problems require a great lot of time and are very expensive. Therefore it is necessary to single out the parts of the structure which need a fracture mechanics analysis. This is made by means of «Fracture control». A «Fracture Control Plan» must be adopted to do it, and to perform the fracture analysis if necessary. «Fracture Control» is a new engineering discipline that interest design, fabrication, environmental control, inspection, maintenance, repair and verification. besides the N.D.I. methods, the fracture control is the new practical aspect of the structure design of today.

The deterministic structure safety approach

While the probabilistic aspects of the structural safety problem were already singled out and the first probabilistic approaches were proposed, the deterministic approach for static loads was improved and brought to a considerably high level of quality so that, together with the fatigue safety concepts, it would constitute a good basis for aeronautics and space purposes. The whole set of such concepts, which were developed when static and fatigue loads seemed to be substantially disjoined from each other, was applied to manned aircraft for many years.

The deterministic approach for static loads is based on the definition of two main load levels: limit loads and ultimate loads. Probabilistic aspects are not absent in their determination but deterministic values are pursued.

Limit loads are the loading conditions which are expected to be encountered in flight and in the other operations within the prescribed flight envelopes.

Ultimate loads are obtained from limit loads by means of a multiplication factor, the «ultimate safety factor», that is usually equal to 1.5 unless otherwise specified.

Roughly speaking structure is requested to sustain ultimate loads, to suffer no unacceptable deformations under limit loads and to suffer no permanent set or yielding after removal of limit loads.

In the previous requirements a more discussed figure was the ultimate safety factor, but the various authors gave different analysis but quite different criticisms. In the commonly adopted philosophy this factor in the early formulation of the approach must give:

— allowance for non-permanent set or yielding at limit load (i.e. must do very close between their requirement at ultimate and at limit load);
— allowance for defects in materials and processing (a more detailed analysis specifies that here undetectable defects are to be intended);
— allowance for design uncertainties and inaccuracies due, for instance, to aeroelastic effects, fatigue, flutter, dynamic effects, structural complexity, loading spectra and load distribution, aerodynamics heating;
— allowance for stiffness;
— allowance for exceeding specified manoeuvres.

During the improvement of the knowledge and technical progress the various uncertainties before indicated were covered by a part requirements or data and that was the reason for several proposals of diminishing the ultimate safety factor. Up-to-day these proposals did not result in practical statements, exception being made for the values lower than 1.5 used missiles and some spacecraft. In their technology are determinant, beside the extreme weight saving necessities, several less severe environmental conditions and very accurate and expressive analysis and test methodologies to cover the more severe environmental conditions.

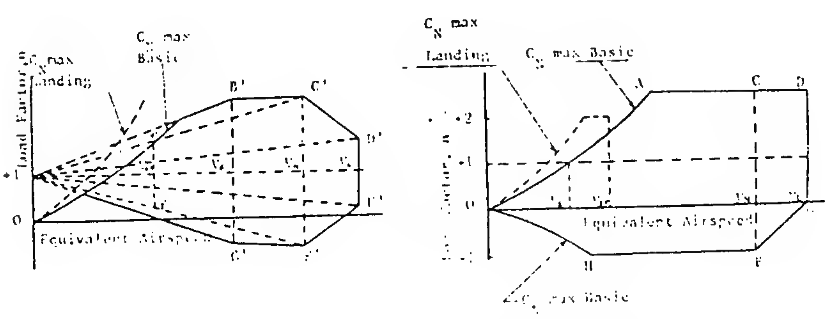


Fig. 2,5 - Manoeuvre and Gust Diagrams valid for Space Shuttle.

Limit load are defined by flight conditions envelopes, such as the manoeuvre diagram and the gust diagram and other specified conditions such as landing, crash, taxing and so on (see fig. 2,5).

Manoeuvre and landing loads have an input on the indication to the pilot as the values of the appropriate parameter not to be exceeded during operational flight are concerned.

Gust loads were derived from statistical data of a phenomenon which cannot be defined in other manners. Gust velocities and the other data on which discrete gust loads are to be evaluated are the result of a choice of the permitted overcome of limit loads in practical operation.

The ultimate safety factor introduces a strong reduction of the hazard of a structural failure.

Other important probabilistic aspects, involved in the so called deterministic approach, are the materials mechanical properties and data which are to be employed during structural calculations and materials evaluation. The data reported in the main handbooks result from statistical evaluations, [5], class «A» data have 99 per cent probability, class «B» data have 90 per cent probability, both with 0.95 confidence level; class «S» data are simply minimum values for acceptance.

As concerns materials an inspection of the consolidated data with a comparison between class A and class B values, seems to indicate also lower variations than the ones reported by Freudenthal [6].

Statistical data and probabilistic structure safety approach

An important change in the design philosophy was the probabilistic analysis of the applied loads. The probabilistic nature of gust loads was clear earlier but, the description available as power spectrum of the root mean square and as function probability of the root mean square to be

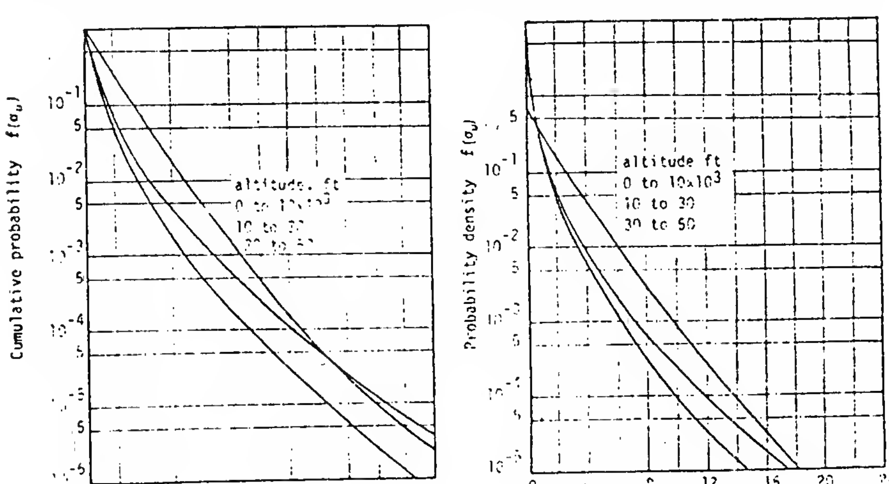


Fig. 2,6 - Cumulative Probability Density of Gusts of given Roof-Mean Square Velocity to be encountered in Flight. [7].

encountered in flight opens new possibility to the structural analysis both as static and fatigue problems are concerned.

Available statistical data on gust power spectrum are in continuous evolution since the early comprehensive data of Press, Meadows and Madlock (see fig. 2,6), [7], to the data to day available and quoted also by the Regulation (see [2]). Also the distribution of the root mean square of the gust velocity (see for instance fig. 4,3, [7]) are continuously under revising process, due to new experimental data acquisition.

Analogous descriptions are today disposable and in continuum improvement as concerns various loading condition as the thrust of missile engines (instability effects), the roughness levels of various airfield preparation (see [2]) the cumulative occurrence of normal load factor experienced at the center of gravity during thousand runway landings (see [2]), the cumulative occurrence of sinking speed during thousand landings and so on.

Recently it was evidentiated that the noise produced by various sources as turbo-jet or other engines, boundary layer and vortices gives not-disregardable acoustic inputs for several parts of missile and aircraft structures (see for instance fig. 2,7). Such inputs create dynamics responses on the structures and fatigue problems as concerns life expectancy, (see also [8]).

Several years ago, the same manoeuvre loads have been analyzed in a statistical manner. The results are given for instance beside the number of peaks in a given time as probability of having peak load factor exceeding given values (fig. 2,8) or the cumulative occurrence of peaks (or through) during a given flight time (see [9]). Such data indicate, as an interesting notation, that there is not zero exceedence of the design manoeuvre load factor due to the pilot manoeuvre action. Also MIL requirements report data on such a probability [2].

Fatigue is also a physical behaviour that implies a statistical description because of the great scatter it is affected by. This scatter is present also in the more simple elementary aspects as, for instance, rotating bending, and depends also on the still unknown nature of the fatigue damage. As long as one considers more complicated loading and geometrical conditions, the scatter becomes greater and greater

The impossibility of having scale model tests and the dependence of the fatigue hebaviour both from general and local geometrical characteristics makes very onerous in time and money to obtain experimental fatigue results on every new structure design.

Therefore the safety phylosophy against fatigue generated two main design concepts, safe life and fail-safe that based mainly on determini-

stic factors, deduced from typical fatigue experimental analysis or practical considerations.

In the safe life concept the structure was requested to suffer no failure in a number of times (deterministic factor that was fixed in 4 the design life time).

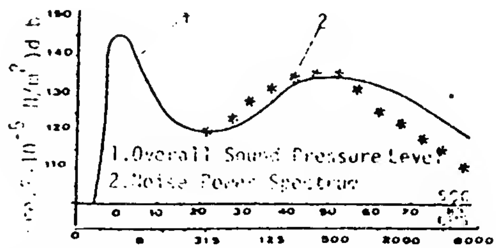


Fig. 2,7 - Acoustic Environment in the Cargo Bay of Space Shuttle [8].

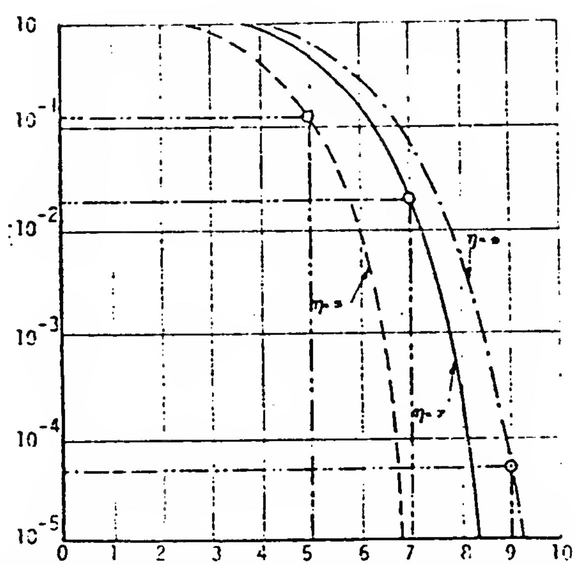


Fig. 2,8 - Probability of exceeding load factor in 'X' hours, from [9].

In the fail safe concept the structure was requested to have redundancies and to suffer no catastrophic failure after the failure of a single element, when subjected to reduced (deterministic factor) static loads in comparison with manoeuvre and gust normal limit loads, so making «not probable» a «catastrophic failure».

Statistical data on the various loading conditions are useful to obtain, after the dynamic analysis (see [10]), statistical data on the appropriate load or stress and, from them, the probability distribution of appropriate quantities (as for instance the peak to trough difference) and their number is a given time of flight. The application of a correlating rule, allows to obtain the life time of the structure before a fatigue failure.

As a sample of statistical properties to be obtained from statistical description of a phenomenon we can consider the case of one gust loads.

The data on atmospheric turbulence are given by means of the two functions below, [11].

1) Probability density f, as function of root mean square gust velocity σ_u

$$f = f(\sigma_u)$$

or as cumulative probability F, as function of σ_u, at various altitude

$$F = F(\sigma_u).$$

2) Power spectral density $\Phi(\Omega)$ of the gust velocity, as function of σ_u and of the reduced frequency $\Omega = \dfrac{2\pi}{\lambda}$ (λ being the wave length),

$$\Phi(\Omega) = \sigma_u^2 \frac{L}{\pi} \cdot \frac{1 + 3\,\Omega^2 L^2}{(1 + \Omega^2 L^2)^2}$$

where $L = 1000$ ft is an acceptable value.

These functions give a probabilistic description of the turbulence that shall be encountered in flight. Their current use is based on the determination of the gust velocity power spectral density $p(\omega)$, where ω is the frequency, taking into account the airspeed, and on the calculation of the power spectral density $p(\omega)$ of every necessary quantity α, as dynamic response of the airplane to the gust input (1).

An asymptotic (for large y) approximate relation between the average number $N(y)$ of maximum per second exceeding a given positive value y and the power spectrum $p(\omega)$ of a stationary Gaussian disturbance, having root mean square σ, is:

$$N(y) = \frac{1}{2\pi} \left[\frac{\int_0^\infty \omega^2 p(\omega)\, d\omega}{\int_0^\infty p(\omega)\, d\omega} \right]^{\frac{1}{2}} e^{\frac{y^2}{2\sigma^2}} \qquad (y > 0)$$

(1) See for instance [10].

If applied also to the α_s' quantities, when there is not the certainty that they are Gaussian (instead the input u is Gaussian), it allows us to obtain an approximate value of the expected cumulative distribution of positive maxima exceeding a fixed positive value y and of negative minima exceeding in modulus a negative fixed value y.

Taking into account the continuous variations in the root-mean square gust velocity, the cumulative distribution of maximum per second exceeding a given positive value y is

$$M(y) = \frac{1}{2\pi}\left[\frac{\int_0^\infty \omega\, p(\omega)\, d\omega}{\int_0^\infty p(\omega)\, d\omega}\right]\int_0^\infty f(\sigma)\, e^{\frac{y^2}{2\sigma}}\, d\sigma.$$

The same expression is valid also for the cumulative distribution of minimum per second exceeding in modulus a given negative value.

Such expression is the basis for the evaluation for instance of the medium number of loading conditions exceeding a given value during an interval of flight time, or the loading condition that can be expected to be exceeded a given number of times in a given interval (for instance between two inspections).

For a discussion of the reliability of the above expression see [12]

Expressions have been evaluated also to obtain approximate evaluations of the fatigue loading conditions. As it si known, fatigue loading is made by cyclic variation of load. The passage from a maximum (minimum) to a minimum (maximum) is an half cyclic variation.

For instance, according to Kowalewski (see [19]), the joint density function of m and α

$$H(m,\alpha) = \frac{N}{\sqrt{2\pi\,\sigma^2(1-I^2)}}\, e^{\frac{m^2}{2\sigma^2(1-I)^2}}\, \frac{\alpha}{\sigma^2 I^2}\, e^{\frac{\alpha^2}{2\sigma^2 I^2}}$$

where m is the mean, α is the amplitude of a certain Gaussian process, and where

N is the number of peak (the number of troughs),
I is the irregularity factor (See [19],
$\bar{N}$ is the zero level crossing with positive slope,

allows us to calculate the environment due to a stationary Gaussian process having a root mean square.

The determination of the power spectral density of load or of a quantity such as the load on an element or the stress at a point that directly connected with the fatigue damage or the requested residual strength, on the basis of the power spectral density of an input such gust, trust instability and so on, is a dynamic response problem.

The classical approach to this problem is the use of the Laplace transform and the determination of the transfer function and in particular of the frequency response that allows us to determinate the response power spectrum.

Recently the dynamic response problem in the case of sonic fatigue, where several hundreds of characteristic modes are involved, have been approached by means of the Statistical Energy Analysis (S.E.A.).

In order to have fully probabilistic approaches of the structural safety problems the difficult is not in the theoretical formulations. In a first analysis it would seem that also the experimental data as the various involved probability distributions do not constitute a problem, but a deeper analysis realized that the significativiness of the statistical figures for the safety is mainly connected with the reliability of those parts of the distribution curves that in a practical formulation can be obtained only by means of arbitrary hypothesis of the type of the probability distributions. At present, this make such figures unreliable. Furtherly, at the present state of the technology, the survival probability of a structural element, in a given flight time is as near to unity that suitable differencies from a case to another are expressed by not self-evident differencies in the correspondent figures.

A schematic representation of the evolution of the thinking on structural safety can be presented by means of some fundamental steps, which can be described as follows.

FAR requirements for transport category airplanes

It is important to note that we are here referring to an old version of FAR Requirements, furtherly subjected to various amendements, in such a manner that to day they are up-dated and take into account the more recent knowledge on the argument.

Such requirements are reported here for their meaning as a step in the evolution of the thinking on structural safety. In my opinion, the safety level of manned space craft would be not lower than the level ob-

tainable with the application, if the passengers are of the same type, of FAR requirements, obviously in the up dailed version.

Beside the usual static requirements, [FAR, vol. 3, Part 25], the quoted old version [11], includes requirements on «Fatigue Evaluation»; that are more interesting for this analysis.

«Those parts of the structure whose failure could result into catastrophic failures of the aeroplane» are requested to be evaluated under provisions of «failure strength», «fail safe strength» and-only for turbojet powered aircraft «sonic fatigue strength».

As concerns fatigue, the structure is requested to withstand the «repeated loads of variable magnitude expected in service» with reference to the «typical loading spectrum expected in service». It is possible to use correctly «service history of airplanes of similar structural design».

As concerns «fail safe» after a failure of a single structural element, «catastrophic failure or excessive deformation» of the remaining structure must be «not probable». This expression is substanzialized by static requirements that the remaining structure must withstand. They are reduced manoeuvre and gust static loading conditions.

Design criteria applied to the NASA space-shuttle structure

Another step in the evolution of the thinking and structural safety is represented by the structural safety criteria prepared for space-shuttle, [12], that explicity mention the probabilistic aspects of material mechanical properties that are indicated also by the up to date handbooks (see, for instance [5]. In particular, material values having a 90% non exceedance probability with 95% of confidence level are required to be used only in «redundant structure in which the failure of a component would result in a safe redistribution of applied loads to other load-carrying members». Materials values having a 95% confidence «may be used whenever failure of a single load path would result in less of structural integrity».

For brittle non-metallic materials «stress level to be used with limit loads shall not permit a probability of more than one failure in a million components».

As concerns the design aspects of fatigue and fracture mechanics, a fail-safe concept is actuated through the requirements that «the failure of a single principal structural eomponent shall not degrade the strength of stiffness of the structure below that necessary to carry a specified percentage of limit load» and that the fatigue life of the remaining structure «shall exceed the time between scheduled inspections».

Safe-life design concepts are required for «all the structure critical to the integrity of the vehicle of personnel safety». The safe-life concept is actuated by requirements including regular inspection for the detection of flaws, and thus it could be better defined as a safe crack-grow-life concept. The flaw which cannot be detected in a regular inspection «should not grow enough before the next scheduled inspection to degrade the strength of the structure below that required to sustain» limit loads at critical temperature.

As concerns «sonic fatigue», cracks and catastrophic failure caused by the cracks themselves must be «not probable», assuming that the loads prescribed for «fail-safe strength», besides several other conditions, «are applied to these areas» which are interested.

In any case, «safe-life» must be «at least four times the specified service life».

Gust problem is approached both as discrete gust requirement and continuous turbulence. As concerns the probabilistic approach, the structure «shall be designed for a 1 per cent or lower risk of exceeding limit loads during the expected time of atmospheric flight».

As concerns safe-life design the so called «safe life tests» are requested. For the design concepts that «depend on non-destructive inspections and flaw-growth predictions» such tests have the purpose of verifying on the structure with artificial flaws «the safe crack-growth predictions» and demonstrating that «non-destructive inspection techniques are adequate». For the design concepts which «depend on non-destructive inspection alone safe-life tests have the purpose of demonstrating «that the techniques are adequate to ensure detection of significant defects».

Up-to-date design criteria (safety requirement for airplane structure damage tolerance)

In order to present the more recent step reference is made to [13], in a way coherente with a conceptual discussion. (For an official use see an original and up dated text). For the sake of brevity the up-date version of FAR requirement is left to the interest of the lecturer.

The last few years made evident the necessity that the airplane safety or flight structure includes among his objectives, [13], «to protect the safety of flight structure from potentially deleterious effects of materials, manufacturing and processing defects through proper material selection and control, control of stress level, use of fracture resistant

concepts, manufacturing and process control and the use of careful inspection procedure».

Safety concepts considered today are indicated below, [13].

— Slow crack-growth structure, where flaws or defects are not allowed to reach the critical size required for unstable propagation.
— Fail-Safe multiple loss path structure, where structure is designed and fabricated in segments which contain localized damage and prevent complete loss of the structure.
— Fail-Safe crack arrest structure, where structure is designed and fabricated in such a way that unstable propagation will be stopped within a continuous area of the structure prior to complete failure.

Degrees of inspectability as considered today are indicated below, together with the corresponding inspection frequency, [13].

— In-flight evident. Once per flight
— Growth evident. Once per flight
— Walkaround Once cvery per flights
— Special visual. once per year or less, if authorized
— Depot or base level, once every one quarter of the design life time. This inspection procedure may include NDI techniques and also foresee removal of components
— In service non-inspectable.

Safety requirements concern the intact structure in case of slow crack-growth structure, and both intact structure and remaining structure after the failure of the critical element in case of failsafe structure.

Two load levels P_{xx} and P_{yy} are introduced in order to give the requirements (see table 2.1).

Tab. 2.1 - P_{XX} loads for various of inspectability, P_{yy} loads (from [13]). (For an official use see an original and up-dated text).

P_{XX^*}	Degree of Inspectability	Typical Inspection Interval	Magnification Factor, M
P_{FE}	In-Flight Evident	One Flight	100
P_{GE}	Ground Evident	One Flight	100
P_{WV}	Walk-Around Visual	Ten Flights	100
P_{SV}	Special Visual	One Year	50
P_{DM}	Depot or Base Level	1/4 Lifetime	20
P_{LT}	Non-Inspectable	One Lifetime	20

* P_{XX}= Maximum average internal member load that will occur once in M times the inspection interval. Where P_{DM} or P_{LT} is determined to be less than the design limit load. The design limit load, shall be the required residual strength load level. P_{XX} need not be greater than 1.2 times the maximum load in one lifetime, if greater than design limit load.

The loads P_{XX}, where xx assumes proper value for each degree of inspectability, are the base for the requirements of the slow crack-growth structure and for the fail-safe structure which remains after a load path failure (or crack arrest).
The load P_{yy} is a minimum load that fail-safe structure must sustain at the instant of load path failure (or crack arrest). P_{yy} shall include a dynamic factor (D.F.) and «should be equal to the internal member load and design limit load or (D.F.) P_{xx} whichever is greater». In the lack of data a D.F. of 1.15 is suggested.

Tab. 2.2 - Slow-crack growth structure (derived from [13]). (Only for non-official use).

Inspectability	Residual strenght req. and damage growth limits for intact structure
In-flight evident	Non-applicable
Ground evident	Non-applicable
Walkaround visual	Non-applicable
Special visual	Non-applicable
Depot or base level	Damage size shall not grow to critical size and shall not cause failure, due to the application of P_{DM}, in a period two times the inspection interval.
In service non-inspectable	Damage size shall not grow to critical size and shall not cause failure due to the application of P_{LT} in a period two times the design lifetime.

Definitions and data for up-to-date design criteria

I - *fracture critical structure*

«Category I fracture critical parts are those components or regions which are sized by the requirements».

«Category II fracture critical parts are those components or regions which could be sized by the requirements if fracture control procedures are not employed».

II - *General requirements*

The safety of flight structure must comply with the requirements at least in a combination of design concepts and inspectability level. Such compliance must be demonstrated by means of all the necessary analytical and experimental work.

All fracture critical regions of all structural components must be inspected as a minimum with a close visual inspection for holes and cutouts and with ultrasonic, penetrant or magnetic inspection for the remainder.

III - *Initial flaw assumptions*

Small imperfections due to material and structure manufacturing and

Tab. 2.3 - Fail-safe multiple load path structure, from [13]. (Only for non-official use).

Inspectability	Residual strength req. and damage growth limits for intact structure	Residual strength req. and damage growth limits for remaining structure subsequent to load path failure
In-flight evident	If structure is depot or base level inspectable for less than failed load path (e.g. subcritical flaws)	Must substain P_{yy} at time of load path failure. 1 shall not cause aircraft failure C P_{FE} during return to base.
Ground evident	a_d shall not grow critical C P_{DM} in one depot or base level inspection interval	Must substain P_{yy} at time of failure. 1 shall not cause aircraft failure C P_{GE} in one flight
Walkarund visual	If structure is not depot or base level inspectable for less than failed load path	Must sustain P_{yy} at time of failure. 1 shall not cause aircraft failure C P_{WV} in 5 times the inspection interval
Special visual	a_i shall not grow to critical C P_{LT} in one lifetime	Must sustain P_{yy} at time failure. 1 shall not cause aircraft failure C P_{SV} in 2 times the inspection interval
Depot or base level		Must sustain P_{yy} at time of failure. 1 shall not cause aircraft failure C P_{DM} in 2 times the inspection interval

a_d = assumed depot or base level damage sizes

a_f = assumed initial flaw sizes

1 = failed load path plus assumed damage in remaining structure

C = due to the application of

Tab. 2.4 - Fail-safe crack arrest structure, from [13]. (Only for non-official use).

Inspectibility	Residual strength req. and damage growth limits for intact structure	Residual strength req. and damage growth limits for remaining structure subsequent to unstable growth and arrest
In-flight evident	If structure is depot or base level inspectable for less than arrested damage (e.g. subcritical flaws)	Must sustain P_{yy} at time of unstable cracking 1 shall not cause A/C failure C P_{FE} during return to base
Ground evident	a_d shall not grow to critical C P_{DM} in one or base level inspection interval	Must sustain P_{yy} at times of unstable cracking 1 shall not cause A/C failure C P_{GE} in one flight
Walkaround visual	or	Must sustain P_{yy} at time of unstable cracking 1 shall not cause A/C failure C P_{wv} in 5 times the inspection interval
Special visual	If structure is not depot or base level inspectable for less than arrested damage	Must sustain P_{yy} at time of unstable cracking 1 shall not cause A/C failure C P_{sv} in 2 times inspection interval
Depot or base level	a_i shall not grow to critical C P_{LT} in one lifetime	Must sustain P_{yy} at time of unstable cracking 1 shall not cause A/C failure C P_{DM} in 2 times inspection interval

a_d = assumed depot in base level damages sizes

1' = damage depending on geometry (see [13])

a_f = assumed initial flaw sizes

C = due to the application of

processing operation (tab. 2.5) shall be assumed to exist in each hole of each element.

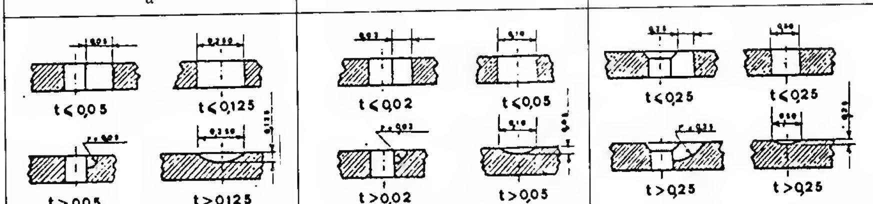
Different (evidently lower) size can be negotiated if the contractor has developed initial quality data on fastener holes.

Besides small imperfections, no more than two initial flaws shall be assumed to exist in any separate element of the structure: one of them

Tab. 2.5 - MIL-A-83444 prescriptions. (Fail safe structures must satisfy further condition not reported here for sake of brevity). (Only for non official use).

Structure characteristic	Inspectability level	Initial length	Inspection interval T_0	Repeated load conditions	Residual strengh load	Crack growth safe interval	Fail safe structure further conditions
Slow crack growth	Depot	see c**	1/4 lifetime	Flight by flight	$P_{DM} = 20 \cdot 1/4$ L.T. M.E.A.L. $P_L \leqslant P_{DM} \leqslant 1{,}2\ P_L$	2 T_0	
Slow crack growth	Non inspectable	see a	1 lifetime	Flight by flight	$P_{LT} = 20$ L.T. M.E.A.L. $P_L \leqslant P_{LT} \leqslant P_e\ P_L$	2 T_0	
Fail safe M.L.P. or C.A.	Depot*	see c***	1/4 lifetime	Flight by flight	P_{DM}	T_0	See MIL-A-83444
Fail safe M.L.P. or C.A.	Non* inspectable	see b	1 lifetime	Flight by flight	P_{LT}	T_0	See MIL-A-83444

a	b	c



Initial lenghts for slow growth structure	Initial lengths for fail safe structures	Crack lengths to be assumed after a depot level inspection

*) For damage of dimensions less than load path or stop length
) Initial length can be assumed as in **a only if the inspection of the component is done with the same NDI procedures used in construction
***) Initial lengths to be assumed after a depot level inspection

in the most critical hole and the other at the most critical location other than a hole. Interaction between them must be considered.

Initial flaw to be assumed depends on design concepts. As concerns slow crack-growth structure, flaw size are specified for hole and cutout loctions, and other locations. Other surface flaw shapes, having the same initial stress intensity factor as the indicated shapes, can be considered as appropriate (see [13], pag. 3).

Small initial flaw sizes other than those specified for slow crack-growth structure may be assumed if it is demonstrated (by an NDI program approved by the procuring activity) that all flaws larger than the assumed

ones have at least a 90 percent. probability of detection with a 95 percent. confidence level.

Smaller initial flaw sizes may be also assumed if proof-test inspection is used, [13], pag. 4.

As concerns fail-safe structures, flaw sizes are specified for hole and cutout locations, and other locations.

Other possible surface flaw shapes with the same initial stress intensity factor (K) shall be considered as appropriate.

IV - *In-service inspection flaw assumptions*

If the component is removed from the assembly and inspected with the same procedure employed during the fabrication, the values indicated as initial flaw assumptions must be used.

When NDI techniques are applied without component or fastener removal, flaw sizes are specified for holes and cutouts locations, and for other locations.

Other possible surface flaw shapes with the same initial stress intensity factor shall be considered as appropriate.

NDI techniques and their impact on safety

Aerospace technologies has developed so far to a very high degree of sophistication, materials and test practices besides design philosophies.

Materials are the result of an industrial process and, the presence of flaws must be realistically accepted.

A flaw can be defined as an inhomogeneity, a discontinuity or a local irregularity at microscopic or macroscopic level.

NDI must reliably provide detection of flaws and of their geometrical location and quantification.

A question may arise if a detected flaw is to be considered a defect, and thus unacceptable. A flaw can incept a crack, and the utmost attention must be paid to the characteristics of fast-growing of small cracks (propagation), and to the previously existing notch effects leading to crack inception.

Of course, the evolution emerging from the quoted requirements was sometimes very fast fas the application, especially when the requirements indicate design criteria which were to be introduced in some technical milieus. Sometime practice came earlier where, besides the indicated safety criteria, design methodologies and criteria have been developed, as for instance when it was necessary to give practical application to the fail-safe concept in its dawning.

In any case, though in the awareness of the differencies between practice and requirements, the analysis of the requirements themselves is clearly indicative of the evaluation of the NDI method impact.

The old version of FAR 25 [11], here considered as the more important statement of a certain step of technological level adopted since 1954, have been successively amended several times. As concerns structural safety, the amendment on «sonic fatigue evaluation» dated 1956, and other amendments on «fail-safe strength», are important to our purpose. NDI methods were not mentioned, and the same «fail-safe» design concept were applied on the basis of the hypothesis of the fatigue failure or obvious partial failure of a single principal element. The principle of periodical inspections and the damage tolerance were not present; moreover neither flaws or cracks were specified as initial imperfection to be taken into account.

Structural design criteria prepared for the application to a space shuttle, [12], dated 1971 (revised 1972), take widely into account the need for new materials, whose data characterization must be made by the same hardware constructor. Among the material characteristics to be evaluated, there are explicitly the material failure mechanisms. The document «is intended as a starting point for preparation of requirements and specifications for the space shuttle». The proof-test is introduced as «non-destructive test» and explicit mention is made of the other NDI methods, in particular with indications on the applicability in the factory and on the field of the various techniques, and with reference to various material types (refractory alloy, ceramics, composites). The development and the evaluation of techniques for inspection are requested. In particular, as concerns safe crack-growth life design concept, it is requested «to demonstrate that non-destructive inspection techniques are adequate». The standards of such NDI methods are the basis for analysis of the flaw-grow because they give «the maximum permissible initial flaw size» to be considered.

When the safe life concepts depend on the proof-test the avoidance or, the reduction of proof-test failures is to be obtained by determining the required NDI «amount and type». The principle of periodic inspection is introduced and the inspection period is the base for the definition of the requirements on fatigue life in the fail-safe concepts and on crack-growth in the safe crack-growth life concept. The great impact of NDI procedures emerges from the statement that, the allowable size of flaws or defects shall be large enough to be detected by practical inspection procedures» and constitute a condition for the adequacy of the design.

The MIL-A-83444, Airplane damage tolerance requirements, [13], dated 1974, represents a strong introduction into the structural safety field of the fracture mechanics analysis and of the periodical inspections principle. They also denote a great confidence into the NDI method capabilities, giving precise indications on the flaws which are to be assumed to exist after factory and in-service inspections. The entire body of requirements is based on «the inspection of 100 percent of all fracture critical regions of all structural components». Flaw sizes lower than the specified sizes may be assumed if «special non-destructive inspection procedures have demonstrated detection capability better than indicated» by the flaw size specifications. The values to be adopted must have at least 90 percent probability to be detected with a 95 percent confidence level. These values must be obtained in a «non-destructive testing demonstration program» «performed by the contractor and approved by the procuring activity».

2.4 Other contributions to hazard in safety

The concepts exposed about structures are applicable also to the structural aspects of the various components different from the structure properly said, as engines, actuator, gear-boxes and so on up to antennas and electronic equipments. When such different components are taken into consideration, failure causes, different from structural ones, take a growing importance following, roughly speaking, the order used in the previous presentation.

In the same order is growing the need of giving probabilistic descriptions of the success capability of each component.

The problem is to obtain reliable data on such success probabilities that could take into account all the failure causes. We will not report an analysis so estended as that indicated in the case of the structural components. It will be sufficient to discuss briefly the allowable sources for the necessary data.

Obviously a suitable and reliable source is given by extensive experimental analysis. In such a case statistical knowledges are necessary in order to schedule the tests and to interpret the experimental results. A general landscape of the involved mathematical knowledges and some typical problems are reported in para. 3.

Another suitable source is given by the observation of the same components if they were installed on former systems that were maintained in exercise.

A last source in order of reliability is the experience previously made

(in particular by the costructor) with similar components, designed and constructed by means of the same technologies.

3. MATHEMATICAL TOOLS FOR PROBABILISTIC EVALUATIONS

3.1 Probabilistic fundamentals

3.1.1 *Probability theory and probability distributions*

Probabilistic concepts usually are divided into «probability theory» calculus and «statistic theory».

Probability calculus is mainly concerned with the determination «a priori» of probabilities of events where complete systems of axioms define the behaviour of the involved components.

Statistical theory is mainly concerned with the (statistical) determination of the properties of populations, that is to say the probability of events that haven't a known dependence on events of known probability.

Probability theory gives us probability distributions based on axioms mathematically formulated. For instence: 1) the probability distribution of a population having a constant (axiomatic) failure ratio, 2) the probability distributions in the case of repeated experiments of known (axiomatic) singular probability, and so on.

It is not in the aims of this paper to review or to exemplify how known probability distributions are obtained in the usual probability theory (see for instance [24] and [26]). Nevertheless it is meaningful to mention that same probability distribution can be obtained with an «a priori» procedure introducing the «uncertainty» concepts and the principle of maximum uncertainty.

The first problem in such a kind of approach is the manner for the evaluation of the uncertainty. If $x_1, \ldots, x_n$ are the values that a variable x can assume, a partial knowledge of the laws regulating the phenomenum can be represented by the probabilities $p_1, \ldots, p_n$ that correspond to the various n values. Following Shannon we can wonder if there exists a

$$H = H(p_1, \ldots, p_n)$$

which can measure in an univocal manner the uncertainty of the probability distribution.

It is noticeable and meaningful that the simplest consistency conditions are sufficient to determine the measure exception made for a multiplication factor. The conditions considered by Shannon are:

1) H must be continuous function fo the p's.
2) When the p's are equal between them $\left(p_i = \frac{1}{n}\right)$ H must be a func tion of n.
3) If groups of the x's and the respective probabilities w's are considered, the uncertainty must satisfy the relation

$$H(p_1, \ldots, p_n) = H(w_1, \ldots, w_r) +$$
$$+ w_1 H\left(\frac{p_1}{w_1}, \ldots, \frac{p_k}{w_1}\right) + \ldots + w_r \left(\frac{p_t}{w_r}, \ldots, \frac{p_n}{w_r}\right)$$

where $w_1 = (p_1 + \ldots + p_k), \ldots, w_r = (p_t + \ldots + p_n)$.

Operating one obtains the result:

$$H(p_1, \ldots, p_n) = -k \Sigma_i p_i l_n p_i$$

where k is a positive constant.

It is now possible to consider the «principle of maximum uncertainty» that is an extension of the Laplace's principle of «insufficient reason». If there exist a known condition of the type

$$\alpha = \sum_{i=1}^{n} p_i f(x_i)$$

beside the obvious

$$\sum_{i=1}^{n} p_i = 1,$$

the principle requires the determination of the p_i's that give the maximum of

$$G = H(p_i) + \mu \sum_{i=1}^{n} p_i f(x_i) + \lambda \sum_{i=1}^{n} p_i$$

where it results $p_i = p_i(\mu, \lambda)$, and the subsequent determination of the constant μ and λ.

Distribution defined in the interval $0 - \infty$ *and having an assigned mean (or expected) value*

If we indicate with v and $p(w)$ the random variable and its probability density and with m the expected value, the maximum uncertainty principle requires

$$\int_0^\infty p(v)\, l_n p(w)\, dv = \text{extremum},$$

with the (integral) conditions:

$$\int_0^\infty v\, p(v)\, dv = m \quad \text{and} \quad \int_0^\infty p(v)\, dv = 1.$$

If $G = p(v)\, l_n p(v) + \mu_1\, v p(v) + \mu_2 p(v)$, the condition $\dfrac{\partial G}{\partial p(v)} = 0$ gives:

$$p(v) = e^{-\mu_1 v - \mu_2 - 1} = e^{-(\mu_2+1)v}\, e^{-\mu_1 v},$$

where μ_1 and μ_2 must be obtained from the two conditions. Operating one obtains:

$$\frac{1}{\mu_1} = m$$

$$p(v) = \frac{1}{m}\, e^{-\frac{1}{m} v}.$$

Therefore, the distribution defined by a given expected value, when the uncertainty is maximized, is the exponential one.

Distribution defined in the interval $-\infty - +\infty$ *and having assigned expected value and variation*

If we indicate with v and $p(v)$ the random variable and its probability density and with $m = E(v)$ and $\delta^2 = \mathrm{VAR}(v)$ respectively the expected value and the mean square deviation, the maximum uncertainty principle requires:

$$\int_{-\infty}^{\infty} p(v)\; l_n p(v)\, dv = \text{extremum},$$

with the conditions:

$$\int_{-\infty}^{\infty} v\, p(v)\, dv = m, \qquad \int_{0}^{\infty} (v-m)^2 p(v)\, dv = \delta^2$$

$$\text{and} \quad \int_{0}^{\infty} p(v)\, dv = 1.$$

If $G = p(v)\, l_n p(v) + \mu_1\, v\, p(v) + \mu_2 (v-m)^2 p(v) + \mu_3\, p(v)$, the condition $\dfrac{\partial G}{\partial p(v)} = 0$ gives.

$$p(v) = e^{-(\mu_3+1)}\, e^{-\mu_2 (v-m)^2}\, e^{-\mu_1 v}$$

where μ_1, μ_2 and μ_3 must be obtained from the three conditions.

The third condition gives:

$$1 = e^{-\mu_3 - 1} \int_{-\infty}^{\infty} e^{-\mu_1 v}\, e^{-\mu_2 (v-m)^2}\, dv\ .$$

Taking into account the previous result, the first condition gives:

$$m = -\frac{1}{2\mu_2} \left\{ e^{-\mu_3 - 1}\, e^{-\mu_2 (v-m)^2}\, e^{-\mu_1 v} \right\} \Big|_{-\infty}^{\infty} - \frac{\mu_1 - \mu_2 2m}{2\,\mu_2},$$

or

$$2\,\mu_2 \mu_1 = \left\{ e^{-\mu_3 - 1}\, e^{-\mu_2 (v-m)^2}\, e^{-\mu_1 v} \right\} \Big|_{-\infty}^{+\infty}.$$

The second member results equal to zero.

Therefore it must be

$$\mu_2 \mu_1 = 0.$$

This means $\mu_1 = 0$ or $\mu_2 = 0$. In the first hypothesis from the second condition, we have,

$$\int_{-\infty}^{+\infty} e^{-(\mu_3+1)}\, (v-m)^2\, e^{-\mu_2 (v-m)^2} = \delta^2.$$

Let us put $V = v - m$, The first member becomes

$$e^{-(\mu_3+1)} \int_{-\infty}^{+\infty} V^2 \, e^{-\mu_2 \cdot V^2} dV.$$

Integrating by parts, by means of some operations, one obtains

$$\frac{1}{2\mu_2} = \delta^2$$

and

$$p(v) = \frac{1}{\sqrt{2\pi}\ v} \, e^{\frac{(v-m)^2}{2\delta^2}} .$$

The case $\mu_2 = 1$ can be disregarded because it doesn't permit to have a mean value m independent from the variation δ (in such a case the distribution would be an exponential one as, it is very easy to verify).

Therefore, the distribution having a given mean value and a given variation is a gaussian one.

The principal probability distributions involved with safety are collected in table 3.1.

A very complex system of analytical connections there exists among such probability distributions in their general or particular forms (see [30]).

Starting from the Binomial distribution B it's possible to obtain the full set of the considered probability distributions exception made for the Γ distribution. On the other hand starting from the Γ distribution it's possible to obtain all the other distributions exception made for B and P distribution. Beside that the «central theorem of statistic» represents a particular connection between each distribution and the «Gauss» one.

In order to have detailed indication it is worthwhile to make reference to specialized books (see for instance Ref. [26].

3.1.2 *Analytical tools and practical assumptions*

The various passages from one to another probability distribution, involving aleatory variable changes and asymptotic processes, need so-

Tab. 3.1 - Present paper used notations. The encircled notations are not used in current bibliography and are introduced only to distinguish among the various probability distributions

Name	Simbol for fig. 1	Variable	Density	Distribution
Binomial	B	k	$b(k)$	
Poisson	P	k	$p(k)$	
Gauss	$N(\mu, \sigma^2)$	x	$f(x)$	$F(x)$
Normal	$N(0, 1)$	u	$\varphi(u)$	$\Phi(x)$
Exponential	E	y	$g(y)$	
Gamma	Γ	z	$h(z)$	
Pearson (or chi-square)	χ_n^2	η	$l_n(\eta)$	
Student (or gosset)	t_n	t	$p_n(t)$	
Fisher	$F_{n,m}$	F	$q_{n,m}(F)$	
Weibull	W	w	$m(w)$	
Rayleigh	R	r	$n(r)$	

me strong analytical tools such as the central theorem of statistic, the generating functions and the characteristic functions.

In order to find theoretical approaches of such analytical tools, reference must be made to specialized bibliography (see for instance Ref. [4] and [5]. We will bound our esposition to a recall of the principal concepts involved in the central theorem of statistic.

Let

$$\xi_1, \xi_2, \xi_3 \dots \xi_n,$$

be a set of independent stochastic variable of which μ_i, σ_i^2, ν_i $(i = 1, 2, 3 \dots n)$ are the moments of the various orders.

Let us consider the variable

$$\zeta^{[n]} = \xi_1 + \xi_2 + \xi_3 \dots \xi_n,$$

which has the moments

$$\mu^{[n]} = \sum_{i=1}^{n} \mu_i; \qquad \sigma^{2\,[n]} = \sum_{i=1}^{n} \sigma_i^2; \qquad \nu^{[n]} = \sum_{i=1}^{n} \nu_i .$$

If

$$\lim_{n \to \infty} \frac{\sqrt[3]{\nu^{[n]}}}{\sigma^{[n]}} = 0$$

the variable

$$u^* = \frac{\zeta^{[n]} - \mu^{[n]}}{\sigma^{[n]}}$$

tends to have the distribution $N(0, 1)$, when $n \to \infty$.

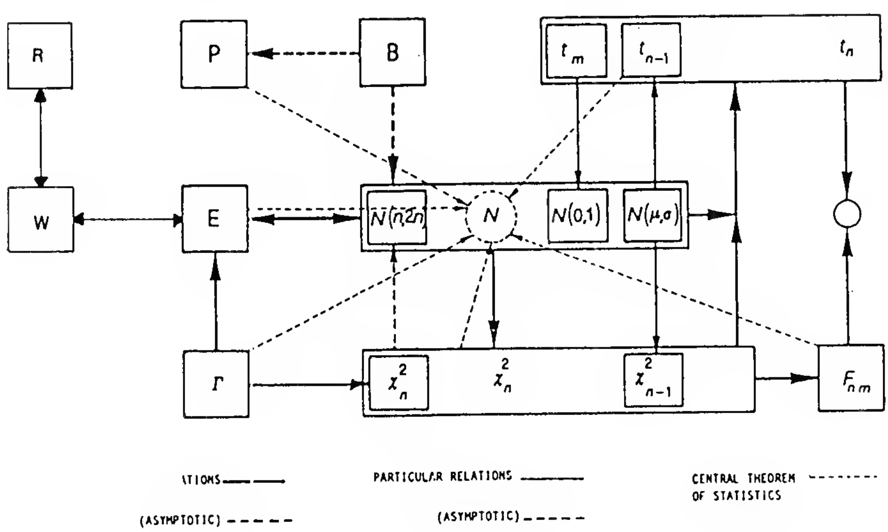


Fig. 3,1 - Probabilistic fundamentals.

Aleatory functions are an aspect of the probability theory which is strongly involved in the structural safety. They are concerned in parti-

cular with the loads encountered in flight. The ergodic theorem has a great usefulness in the handling of the stationary aleatory functions. It is an analogue of the laws of large number and states that a stationary aleatory function $(u(t))$ and the product

$$u(t+\tau)\,u(t)$$

have mean values where no matter the means are made for respect to the time or at a given time for respect to the probability field.

The applications contained in this paper don't make use of the ergodic theorem, due to the particular choice of the applications themselves. Therefore, it is sufficient to have recalled such theorem in order to emphasize the strong impact that aleatory functions have in a complete review of the probabilistic aspects of structural safety.

In the practical applications it is worthwhile to make use of practical assumptions that allow to approximate by means of disposable values, other values that would require time and efforts for their calculation. The introduction of the electronic elaborators will reduce such kind of approach. In any case it is useful to report an example of practical approximation, in order to emphasize that the assumption are always based on theoretical results. As an example we can make reference to the connection illustrated in fig. 10, where the chi-square distribution of n degrees of freedom is quoted as the distribution of an aleatory variable involving n values, kept-off from a population having a gaussian distribution.

The chi-square tables usually indicate data for $f=n-1<30$. Since chi-square distribution tends to the Gauss' one when $n\to\infty$, in this case it is

$$\mu=n-1, \quad \sigma=\sqrt{2(n-1)}, \qquad \text{(see 13)},$$

and it is possible to use the aleatory variable

$$u=\frac{\eta-(n-1)}{\sqrt{2(n-1)}}$$

which has a normal distribution.

The variable $\sqrt{2\eta}=x$ has asymptotically $(n\to\infty)$ the distribution $N(\sqrt{2n-3},\ 1)$.

The variable $u=\sqrt{2\eta}-\sqrt{2n-3}$ has asymptotically $(n\to\infty)$ the distribution $N(0,\ 1)$.

Let ξ be an aleatory variable, having mean value μ_ξ and mean square σ_ξ^2. In several technical areas it is necessary to evaluate the probability that $\xi \gtrless \xi_0$, where ξ_0 is a fixed value, or the probability that $\xi_1 < \xi < \xi_2$ where ξ_1, and ξ_2 are fixed values, or the values ξ_a and ξ_b ($\xi_a < \xi_b$) for which $\xi_a < \xi < \xi_b$ with a wanted probability.

The same necessity can exist as concerns μ_ξ, σ_ξ^2 other quantities to the probability distribution.

The problem can be approached experimentally through an (approximated) evaluation based on mean values or equivalent operations (see for instance ξ and a subsequent statistical analysis of the results. The aspect that must be statistically examined is the «probability» that the approximated evaluation b_e of the wanted value.

The knowledge of the statistic distribution of the difference between (experimental) approximated and true value, when the «sample» is repeated an infinite number of time. (In this case the «sample» is the set of the experimental results on which the approximated evaluation is based). As typical problem, let x be a variable having gaussian distribution $N(\mu, \sigma)$. Let us consider a sample of a degree of freedom picked-out from the population x.

The experimental approximated mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

has the gaussian distribution $N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$. If σ is known, the variable

$$u - \frac{\bar{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$$

has the normal distribution.

Therefore

$$P(u \leqslant u_1) = \phi_1(u) = \int_{-\infty}^{u_1} \varphi(u)\, du$$

and

$$P(u \geqslant u_1) = 1 - \phi_1(u) = \int_{u_1}^{\infty} \varphi(u)\, du.$$

Hence, for instance,

$$P\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leqslant u_1\right) = \Phi_1(u)$$

$$P\left(-\mu \leqslant u_1 \frac{\sigma_n}{\sqrt{n}} \bar{x}\right) = P\left(\mu \geqslant \bar{x} - \frac{\sigma_n}{\sqrt{n}}\right) = \phi_1(u)$$

and, chosen a value μ_1, $P(\mu \geqslant \mu_1)$ can be obtained, as follows:

$$\mu_1 = \bar{x} - u_1 \frac{\sigma_n}{\sqrt{n}}$$

$$u_1 = \frac{\sqrt{n}}{\sigma_n}(\bar{x} - \mu_1).$$

From u_1, one can obtain ϕ_1 by means of eq. 1).

As another example, chosen μ_1 and μ_2 ($\mu_1 < \mu_2$), the probability that μ may be within $\mu_1 - \mu_2$ is

$$P(\mu_1 \leqslant \mu \leqslant \mu_2) = \phi_1 - \phi_2.$$

where ϕ_1 and ϕ_2 correspond to u_1 and u_2 and u_2 and u_2 are obtained from eq. 2).

From a different point of view, chosen a probability α (confidence level), the simmetric or unilateral (confidence) interval can be determined, where with the chosen «confidence level» μ may be included.

bilateral interval

$$-u_{\alpha/2} \leqslant \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leqslant u_{\alpha/2}$$

$$\bar{x} - u_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{x} + u_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

unilateral interval

$$-u_\alpha \leqslant \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

$$\mu \geqslant \bar{x} - u_\alpha \frac{\sigma}{\sqrt{u}}.$$

3.2 Applications

3.2.1 *Interpretation of NDI methods capability tests*

The tests are usually made by means of a number of operators, which make observations on g groups of $N_p \geqslant n_p$ specimens ($p = 1, 2 \dots g$), where n_p of them include a crack of length l_p.

The tests are based on the assumption that to each value l_p and with defined test condition, there exists a probability p_p of detecting the crack included in the specimen. The scope of the tests is the statistical determination on a sample of the probability p_p, distinguishing or not between an operator and another.

Since the n_p cracked specimens belong to a group of N_p specimens, there are the following possibility:

1) cracked specimen	detected as cracked
2) non cracked specimen	not detected
3) cracked specimen	not detected
4) non cracked specimen	detected as cracked.

Usual interpretation consider the results of 1) and 3), regarding n_p cracked specimens, in order to evaluate p_p and separately the results of 2) and 4), regarding $N_p - n_p$ specimens, in order to evaluate the probability of detecting as cracked a non cracked specimen. Clearly the proper theoretical analysis is that of «repeated tests».

As it is well known, the «repeated tests» follow the binomial distribution, having the density

$$B_k = \binom{n}{k} p^k q^{n-k}$$

where k is the stochastic variable and B_k is the Probability that of n tests k have positive results.

The purpose of the experimental tests is to obtain a statistical evaluation of p. Beside the variable k one can consider also the frequency k/n.

Let $\bar{k}$ be the number of favourable results which is obtained experimentally. (See also table 3.2).

Tab. 3.2 - Estimating variable for some statistical quantities related to the binomial distribution

Variable	Mean	Estimating variable	Variance	Estimating variable
k	$\bar{k}=np$	$\bar{k}$	$\sigma_k^2=npq$	$s_k^2=\dfrac{\bar{k}(n-\bar{k})}{n-1}$
$v=\dfrac{k}{n}$	$\bar{v}=\dfrac{\bar{k}}{n}=p$	$\dfrac{\bar{k}}{n}$	$\sigma_{k/n}^2=\dfrac{pq}{n}$	$s_{k/n}^2=\dfrac{\bar{k}(n-\bar{k})}{n^2(n-1)}$

If $\bar{k}\,(\bar{k}/n)$ is adopted as estimating variable for $n_p(p)$, when n is great, one can admit that the variable

$$\frac{\bar{k}-np}{\sigma_k/\sqrt{n}}\;\left(\frac{\bar{k}/n-p}{\sigma_k/\sqrt{n}}\right)$$

has a normal distribution. Therefore

$$P\left(\left|\frac{\bar{k}}{n}-p\right|\leqslant r\sqrt{\frac{p(1-p)}{n}}\right)=\int_{-r}^{r}\varphi(u)du=\Phi(-r)-\Phi(r);$$

$$\left(\frac{\bar{k}}{n}-p\right)^2\leqslant r^2\frac{p(1-p)}{n}.$$

The equation in the unknown p which results by considering the sign of equality, has the roots

$$\left.\begin{matrix}P_1\\ \\ P_2\end{matrix}\right\}=\frac{1}{2(n+r^2)}\left[2\bar{k}+r^2\mp r\sqrt{r^2+4\bar{k}\left(1-\frac{\bar{k}}{n}\right)}\right];$$

such

$$P(p_1 < p < p_2) = \int_{-r}^{r} \varphi(u)\, du.$$

Actually the specimens examined are N_p. The probability of correctly detecting a cracked specimen in each test is

$$p^* = \frac{n_p}{N_p}$$

composed by a known factor n_p/N_p and the quantity p, object of the experimental analysis. What has been said for the sample n_p n order to determine p, can be repeated for the sample N_p in order to determine p^*. One obtains

$$P(p_1^* < p^* < p_2^{*\prime}) = \int_{-r}^{r} \varphi(u)\, du =$$

$$= P\left(p_1^* \frac{N_p}{n_p} < p < p_2^* \frac{N_p}{n_p}\right).$$

In this procedure, by changing the actual difference $N_p - n_p$, one can experimentally analyze the influence on the probability p of the dimension N_p.

A similar analysis can be made in order to obtain a statistic evaluation of the probability of detecting one of the specimens, in a population of N_p specimens.

The described procedure is suitable for high values of n (for instance $n > 40$), where the approximation of the binomial distribution by means of the Gauss' one give no practical differences.

In the NDI method practice it is still necessary to evaluate the method capabilities by means of procedures which can keep their practical validity also for very low values of n. In several experimental result evaluations a procedure has been adopted which, doing still reference to the Gauss probability distribution, evaluates the standard deviation of such a distribution in the following manner. Reference is made to the observation that the aleatory variable

$$\eta = \frac{(n-1)\bar{x}}{\sigma_x^2}$$

(x correspond in the continuum field to the discrete k) has a distribution near to χ^2 of $n-1$ degrees of freedom. Thus a confidence limit for

$$\sigma_x^2 = \frac{(n-1)\bar{x}}{\eta}$$

can be obtained corresponding to each wanted value of the confidence level. If $n < 40$, the confidence limit of σ_x can be also obtained from Gauss' distribution, but adopting the procedure of the stochastic evaluation of σ_x, in order to have no discrepancies between high and low values of n,

When the binomial distribution is approximated by the Gauss' one — the more n is great the more such an approximation is valid — the probability distribution becomes:

$$P_x = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_x} e^{-\frac{1}{2}z^2}$$

$$\left(z = \frac{x-\hat{x}}{\sigma_x};\ \hat{x} = np\right).$$

If σ_x is known through an independent way, one can evaluate the externum x_0 of the unilateral confidence interval for which $p > x_0/n$ with the probability $1-\alpha$ as follows. Let $x_{-2\alpha}$ and $x_{2\alpha}$ be such that

$$\int_{-x_2}^{x_2} P_x\,dx = 1-2\alpha, \quad |x_{2\alpha} - np| = b_{2\alpha}\sigma_x \qquad (b_{2\alpha} > 0)$$

it also

$$P(|\bar{x} - np| < b_{2\alpha}\sigma_x) = 1-2\alpha.$$

Actually $n \geqslant 0$, $p \geqslant 0$, $\bar{x} \geqslant 0$. This

$$P[(np + b_{2\alpha}) > \bar{x}] = 1-\alpha$$

$$P[np > \bar{x} - b_{2\alpha}] = 1-\alpha.$$

The next step is now to put in this relation a value σ_x^e of σ_x, stochastically obtained with wanted $1-\beta$ «confidence level».

If η_β is the value of η where in the X^2 distribution of $n-1$ degrees of freedom

$$P(\eta > \eta_\beta) = 1-\beta$$

we have

$$P\left[\frac{(n-1)\bar{x}}{\sigma_x^2} > \eta_\beta\right] = 1-\beta = P\left[\frac{(n-1)\bar{x}}{\eta_\beta} > \sigma_x^2\right]$$

and

$$P\left[\sigma_x^2 < \frac{(n-1)\bar{x}}{\eta_\beta}\right] = 1-\beta.$$

The value

$$\sigma_{x,\beta} = \frac{(n-1)\bar{x}}{\eta\beta}$$

is the extremum of the unilateral confidence interval for lower values of σ_x, with the «confidence level» $1-\beta$. The value σ_x^* is such that

$$np > \bar{x} - b_{2\alpha}\,\sigma_{x,\beta}$$

is $1-\alpha$ probable with $1-\beta$ confidence level in the stochastic evaluation of σ_x.

Therefore, it is possible to use relation 10″, rigorously valid for Gauss' distributions and $n \to \infty$.

Are now comprehensible the NDI method capabilities which in function of the dimension are quoted as: probability of detection (stochastic evaluation of p) for a given probability (evaluation of σ) and of a given «confidence level» (evaluation of p).

3.2.2 *Probabilistic interpretation of safety in static load failure*'

If we indicate with A and F respectively the applied loads and the strength and with $g(A)$ and $f(F)$ respectively their probability densities, the probability of survival (or safety) when a single load is applied to a single structure can be calculated as

$$S_1 = \int_{-\infty}^{\infty} f(F)\, dF \int_{-\infty}^{F} g(A)\, dA$$

where a sequence of a number of n separated load each belonging to A is applied to the same structure, the survival of the structure needs the survival under each of the n applied loads. From a logic point of view the n loading applications are connected as a series of n operation having the same safety S_1 and the total safety can be expressed by means of the expression:

$$S_n = S_1^n .$$

Putting now $-\bar{\lambda} = l_n S_1$ we easily obtain:

$$S_n = e^{-\bar{\lambda} n}, \ (\bar{\lambda} \text{ results positive because } S_1 \leqslant 1),$$

or, in other words, an exponential expression of the cumulative probability of survival after a sequence of n applied loads.

If we now admit that there exists a linear relation between time t and the number n of the applied load,

$$t = an,$$

the previous expression becomes $(\lambda = \bar{a}\bar{\lambda})/0$

$$S(t) = e^{-\lambda t}.$$

The cumulative failure probability is, obviously,

$$1 - S = 1 - e^{-\lambda t},$$

and the failure probability density is

$$f(t) = \lambda e^{-\lambda t}.$$

In other words, the admission of a constant distribution of the loads during the time ($t = an$) and the hypothesis that the various loads of the sequence belong to the same population A, determine a behaviour model with a constant failure rate.

Approximate expressions of the cumulative survival probability can be derived observing that when S_1 is very close to one, it is:

$$l_n\, S_1 \cong S_1 - 1.$$

Then

$$-\bar{\lambda} \cong S_1 - 1, \quad \lambda = a(1 - S_1)$$

and

$$S(t) = e^{-a(1-S_1)t}.$$

Furtherly, when $a(1 - S_1)t$ is very close to zero, the approximate expression

$$S(t) = 1 - a(1 - S_1)t$$

is valid.

3.2.3 *Influence of the loaded zone extension on the fatigue behaviour*

The aim of this paragraph is to give a probability theory insight on the well known «scale effect», where a larger scale specimen, subjected to identical (in the corresponding points) stress levels, has lower fatigue number of cicles.

The considerations that follow are referred to cylindrical beams in tension in order to correlate between them the fatigue behaviour of specimens of different length also the effect of the reduction of the fatigue number of cycles with the increase of the specimen length is well known. An extension of the same considerations to the proper «scale effect» is not difficult.

Let us consider a constant section specimen loaded in tension, whose fatigue initial crack probability density $d_1(n)$ have been determinated through an experimental analysis under well defined test conditions (L, $N(t)$, and so on).

A question arises; what is the probability that a crack can start in a specimen of length 2L?

It is as if samples of 2 specimens of length L are drown and for each of such samples the lowest t is considered;

$$p_2(n_2) = 2p_1(n_2) \int_{n_2}^{\infty} p_1(n_1)\, dn_1 .$$

In the case of length jl we have:

$$d_j(n_j) = jd_1(n_j) \int_{n_j}^{\infty} (j-1)\; p_1(n_{j-1})\, dn_{n-1} \ldots$$

$$\ldots \int_{n_{j-i+1}}^{\infty} (j-i)\; p_1(n_{j-i})\, dn_{j-i} \ldots \int_{n_2}^{\infty} p_1(n_1)\, dn_1 .$$

The application of the previous consideration to an exponential distribution (see fig. 19a) and to a Gauss' distribution (see fig. 19b) gives a (qualitative) indication of the studied effect.

3.2.4 *Some statistical considerations on the validity of the Miner's rule*

The aim of this paragraph is to dicuss some statistical aspects of the rule proposed by Miner (and independently by Palmgren). Such rule, as it is well known, supposes that the fatigue rupture is reached when

$$\sum_{i=1}^{n} \frac{\Delta n_i}{N_i} = 1 \qquad (i = 1,\ 2,\ 3 \ldots\ n)$$

where Δn_i is the number of cycles of i-type and N_i is the corresponding rupture number of cycles inconstant amplitude tests.

In the previous statement N_i is managed as it were a deterministic value associated with a particular constant amplitude loading cycle. Beside that, in the applications no attemption is made to the sequence of application of the various loading cycles.

At the present state of art the validity of the Miner's rule, is under discussion, mainly in connection with the influence of the loading cycle sequence. This is particularly important during the crack propagation and in the case of ductile behaviour.

This paragraph proposes some considerations on the non deterministic nature of cycles, making reference to the case where the sequence of the loading cycles is not important.

From an engineering point of view, the typical problem to be considered is the determination of the probability distribution of the rupture

number of cycles when in the loading sequence there is the probability (p) of having «type-a» cycles and the probability $(1-p)$ of having «type-b» cycles. This on the base of the knowledge of the probability distributions of the rupture number of constant amplitude type-a and type-b cycles.

Let us make two main basic assumptions

I) the number n of cycles of each type is so large that it si possible to assume that also in an increment Δn there are $p\Delta n$ type-a cycles and $(1-p)$ type-b cycles.

II) A loading cycle produces a failure rate which only depends for each cycle type on the actual survival function, even if the amplitude of the loading cycle sequence is not constant.

If the failure rate of a generic distribution is

$$\lambda(n)=\frac{f(n)}{R(n)}, \qquad \left(R_n=1-\int_{\infty}^{n} f(n)\, dn\right)$$

the wanted failure rate is

$$\lambda(n)=p\frac{f_a(R)}{R}+(1-p)\frac{f_b(R)}{R}$$

and the correspondent density of probability is

$$f(n)=p\, f_a[t_a(R)]+(1-p)\, f_b[t_b(R)].$$

In this expression $t_a(R)$ and $t_b(R)$ are the probability densities in function of R, respectively for the type-a and type-b constant amplitude cycles.

Let us now suppose that the following conditions are verified.

a) Each of the two distribution density *fa* or *fb* is depending only on a parameter λ_a or λ_b.

b) The two density $f_a(\lambda_{a,b})$ and $f_b(\lambda_{b,n})$ have the same analitical expression, exception made for the difference between λ_a and λ_b (es.

$$f_a=\lambda_a e^{-\lambda_{an}}; \quad f_b=\lambda_b e^{\lambda_{bn}}.$$

Therefore, the expected values have the same analytieal expression
c) It is possible to put

$$f_a[n(R)] = \alpha(\lambda_a)\cdot\beta(R), \quad f_b[n(R)] = \alpha(\lambda_b)\cdot\beta(R)$$

$$\mu_a = E[\alpha(\lambda_a)], \quad \mu_b = E[\alpha(\lambda_b)].$$

If these three conditions are verified, one obtains

$$\frac{f(n)}{\beta(R)} = p\alpha(\lambda_a) + (1-p)\,\alpha(\lambda_b).$$

The second member is independent from n and R. Thus the wanted $f(n)$ has the same analitical expression of f_a and f_b, with

$$\alpha(\lambda) = p\alpha(\lambda_a) + (1-p)\alpha(\lambda_b).$$

The expeeted value

$$\mu = E[\alpha(\lambda)] = E\{p\alpha(\lambda_a) + (1-p)\alpha(\lambda_b)\}$$

can now be related to μ_a and μ_b. If the further condition that

$$E[\alpha(\lambda)]$$

is proportional to $1/\alpha(\lambda)$ is verified, the Miner's rule

$$\frac{1}{\dfrac{\pi\mu}{\mu_a} + \dfrac{(1-p)\mu}{\mu_b}} = 1$$

holds.

As a sample we can consider the case of exponential distribution:

$$f_a(n) = \lambda_a e^{-\lambda_a n} \quad f_b(n) = \lambda_b e^{-\lambda_b n}$$

$$R_a = e^{-\lambda_a n} \quad R_b = e^{-\lambda_b n}$$

$$\dot{n}_a(R) = -\frac{ln\,R}{\lambda_a} \quad n_b(R) = -\frac{ln\,R'}{\lambda_b}$$

$$f_a[n(R)] = \lambda_a R \quad f_b[n(R)] = \lambda_b R$$

$$\alpha(\lambda_a)\lambda_a;\ \beta(R) = R \quad \alpha(\lambda_b) = \lambda_b;\ \beta(R) = R$$

$$\mu_a = E\alpha(\lambda) = \frac{1}{\alpha(\lambda_a} = \frac{1}{\lambda_a}; \quad \mu_b = \frac{1}{\lambda_b}$$

$$\frac{f(n)}{\beta(R)} = \frac{f(n)}{R} = p\lambda_a + (1-p)\lambda_b$$

$$\mu = \frac{1}{\lambda} = \frac{1}{p\lambda_a + (1-p)\lambda_b} = \frac{1}{\dfrac{p}{\mu_a} + \dfrac{1}{\mu_b}}$$

$$\frac{1}{\dfrac{p\mu}{\mu_a} + \dfrac{(1-p)\mu}{\mu_b}} = 1.$$

This means that the mean value μ of the rupture number of cycles satisfies the Miner's rule.

Thus we can conclude that under the two basic assumptions I) an II), the conditions a), b), c) and d) are sufficient for the validity rule.

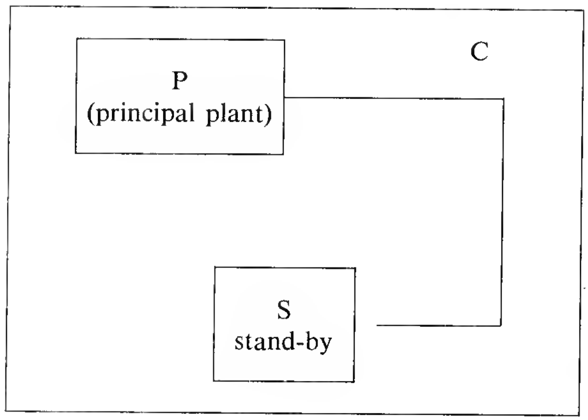
The question if such conditions are also necessary is not examined in this work as well as what happens under basic assumption different from I) and II).

3.2.5 *Stand-by of an auxiliary plant*

The problem is the determination of the failure probability of the complete system C (see Fig. 3.2) when the failure probability of the components are known. If P and S have constant failure rate λ_P *and* λ_S respectively, $f_P(t)\,dt$ being the probability of failure of P in dt; we have:

$$f_P(t) = \lambda_P\, e^{-\lambda_p t} \quad f_S(t) = \lambda_S\, e^{-\lambda_S t}.$$

Fig. 3.2 - Scheme of a system with a stand-by.



The probability that C has a failure before t_0 when P had a failure in dt after t $(t_0 > t)$ is:

$$f_P(t)\,dt\,\{1 - e^{-\lambda_S(t_0 - t)}\}.$$

If $G(t_0)$ is the probability that C a failure before t_0 is

$$G(t_0) = \int_0^{t_0} f_P(t)\,dt\,\{1 - e^{-\lambda_S(t_0 - t)}\}.$$

Operating one obtains:

$$G(t_0) = 1 - e^{-\lambda_P t_0} - \frac{\lambda_P}{\lambda_P - \lambda_S}\{e^{-\lambda_S t_0} - e^{-\lambda_P t_0}.$$

If $\lambda_S \to \infty$ (non operative stand-by), $G(t_0) \to 1 - e^{-\lambda_P t_0}$, obviously.
If $\lambda_S \to \lambda_P$, we obtain:

$$[G(t_0)]_{\lambda_S = \lambda_P} = 1 - (1 - \lambda_P t_0)\, e^{-\lambda_P t_0}.$$

To the same result we can arrive using the Poisson distribution. In fact the probability of having not any failure in to is a $e^{-\lambda_P t_0}$. The probability of only one failure (of the P component, of course) is λ_P to

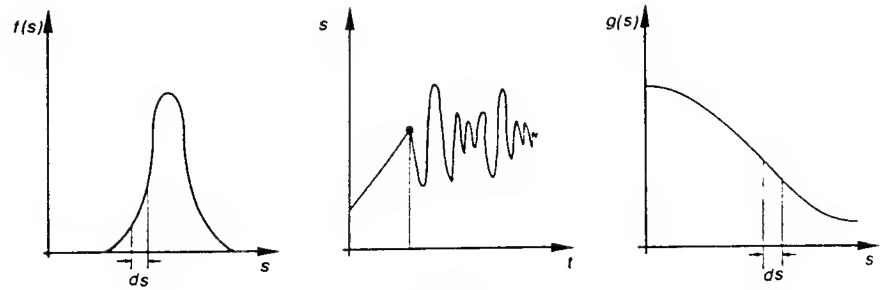


Fig. 3,3

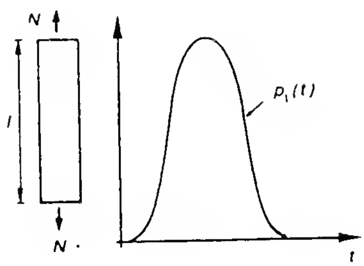
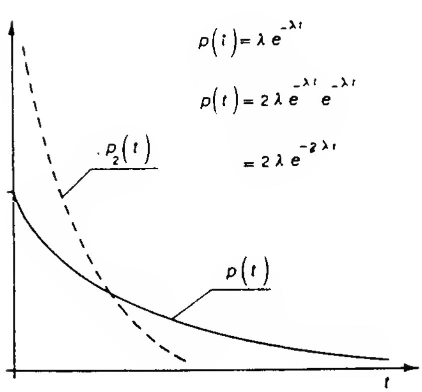


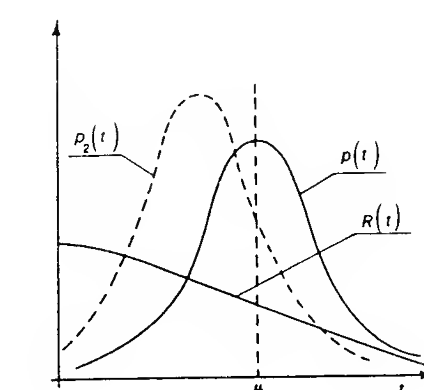
Fig. 3,4 a)

Fig. 3,4 b)

Fig. 3,5 a)

Fig. 3,5 b)





$e^{-\lambda_P t_0}$. Thus, the probability of a correct functioning in t_o is $e^{-\lambda_P t_0}+$ $+\lambda_P t_0 e^{-\lambda_P t_0}$, and the probability of failure is

$$G(t_0)=1-(1+\lambda_P t_0)\ e^{-\lambda_P t_0},$$

as previously obtained.

Such a result is based on the observation that the event of the correct functioning in t_0 can be considered as the event of having only 0 or 1 failure in $u \to \infty$ interval dt, when there is the probability of failure q such that $nq=\lambda_P t_0$. The probability **q** being offered by P in $0 \vdash t$ and by P or S in $t \vdash t_0$.

4. CONCLUSIONS ON SAFETY AND FUTURE ASTRONAUTIC ACTIVITIES

The manned space activity can be considered as the natural sofisticated extension of the aeronautical activity — in any case it was actually developed in such a way.

Therefore, a menned space mission would be subjected to the design condition of having a demonstrated predetermined overall safety level, taking in to account all the hazard sources.

Obviously, the wanted safety levels can take into account the particular nature of each single space mission and the kind and the training of the crew and passengers.

In order to contribute to discuss how to obtain, in our quality of scientists and engineers, aceptable safety levels for future astronautic activities a general survey of concepts and mathematical tools have been presented.

If we consider the safety determined by the internationally accepted «Airworthiness Standards» for Transport Category Airplanes as a sample overall level of safety and the safety obtained in the past in the experimental manned space activity as another sample level, it is my opinion that in space flight the safety must be near as close as possible, to that of the corresponding aeronautical activity.

It is well known that the overall safety levels are usually determined by a very complex system of safety criteria and precautions, that are coherent with the various design problems and that participate as necessary (or convenient) both of deterministic and probabilistic nature. The overall safety level is of very hard evaluation, in particular as structural safety is concerned — on the other hand structural safety level is usual-

ly so high that its quantization in practice is not necessary. Such considerations will be applicable also to the future activity.

Therefore, the problem is to obtain a complex system of criteria and prescriptions, strictly connected with the particular life and mission conditions, that could reasonably assure an acceptable overall safety level.

REFERENCES

[1] MIL-STD-721 - Definition of effectiveness terms for reliability, maintenability, human factors and safety.

[2] MIL-A-8866B (ASG) - Airplane strength and rigidity reliability requirements, repeated loads and fatigue. 1975.

[3] HOFF N.J., *Philosophy of safety in the supersonic age*. AGARD Report 87.

[4] EBNER H., *The problem of structural safety with particular reference to safety requirements*. AGARD Report 150.

[5] MIL-HDBK-5A - Metallic materials and elements for aerospace vehicle structures.

[6] FREUDENTHAL A.M., *Safety and safety factors for airplanes*. AGARD Report 153.

[7] PRESS H., MEADOWS M.T., HADLOCK I., *A reevaluation of data on atmospheric turbulence and airplane gust loads for application in special calculations*. NASA Report 1272.

[8] ANTONA E., GIAVOTTO V., SALVETTI A., VALLERANI E., *The role of fracture mechanics and fatigue in the design of advanced aerospace vehicle*. ICAS Paper n. 76-26.

[9] MANGURIAN G.N., *The aircraft structural factor of safety*. AGARD Report 154.

[10] ETKIN B., *Dynamics of atmospheric flight*. Ed. J. Wiley & Sons.

[11] F.A.R. Transport Category Airplanes. Vol. III, Part 25, 1969.

[12] NASA - Space Vehicle Design Criteria Structural Design Criteria Applicable to Space Shuttle, NASA SP-8057. Rev. March 1972.

[13] MIL-A-83444 (USAF) - Military Specification. Airplane damage tolerance requirement. 2 July 1974.

[14] Bolis E., *Non destructive Inspection Practices.* AGARDograph n. 201, Voll. I e II.

[15] AERITALIA S.p.A. - Evaluation of crack detection methods. Phase I Report, Sept. 1976.

[16] Rummel W., Todd P., RathkeR. & Castner W., *The detection of fatigue cracks by nondestructive test methods.* Materials evaluation, october 1974.

[17] Cooper T., Packman F. & Yee G. eds., *Prevention of structural failure. The role of quantitative nondestructive evaluation.* American Society for Metals, Ohio, 1975.

[18] Liebowitz H. Ed., *Fracture mechanics of aircraft structures.* AGARD-AG-176.

[19] Huck M., Schutz W., Fischer R. & Koller H.G., *A standard random load sequence of Gaussian type recommended for general application in fatigue testing.* LBF Report 2909, IABG Report 570, Aprile 1976.

[20] Antona E., Giavotto V., Salvetti A., Vallerani E., *Meccanica della frattura in strutture spaziali pressurizzate e soggette a fatica acustica.* IV Congresso Nazionale A.I.D.A.A., Milano, 19-23 Settembre 1977.

[21] Gay J., *Recherche experimentale de la tenue des structures d'avion à la fatigue acoustique.* AGARD-CP-113.

[22] Antona E., *Critical review of various structural safety concepts taking into account NDI methods,* in AGARD, C.P. n. 234.

[23] Eggwertz S. and Lindsjo G., *Influence of detected crack length at inspections on probability of fatigue failure of wing panels.* ICAF, doc. n. 834.

[24] Papoulis, *Probability, random variables and stochastic processes.* McGraw-Hill, Kogakusha, 1965.

[25] Bussetti G., *Esercitazioni di Fisica.* Ed. Levrotto & Bella, Torino, 1974.

[26] Galeotti P., *Calcolo delle probabilità e statistica.* Ed. Levrotto & Bella, 1976.

[27] Pikman P.F., Klima S.J., Davies R.L., Malpani J., Moyzis J., Walker W., Yee B.G.W. and Johnson D.P., *Reliability of Flaw Detection by Nondestructive Inspection.* Metals and Book. 8ª Ed. - Vol. 11 - American Society for Metals.

[28] AGARD (Several Authors) - Factors Safety, AGARD Report No. 661, November 1977.

[29] FIORINI V., *Problemi di aeronavigabilità degli impianti*, Atti del Siminario sui sistemi aeronautici e spaziali, Torino 5-6 Giugno 1984, Ed. Levrotto & Bella, Torino.

[30] ANTONA E., *Mathematical aspects of the probability evaluation of structural safety and NDI capability.*

[31] ANTONA E., *Sicurezza, durata e affidabilità delle strutture aerospaziali*, Notiziario AMMA, Torino, Maggio 982.

[32] LEPORATI-LEVI, *Principes de Sécurité Structurelle Appliquée au Bétan Armé*, C.M.E.C, Paris, 981.

[33] LEPORATI, *The assessement of Structural Safety,* Research Study Prss, Letcworth, Hearts England.

[34] ANTONA E., *Probability Principle and Static Load Safety*, Atti della Accademia delle Scienze di Torino, (to be published, presented Dicember 1987).

[35] ANTONA E., *Sicurezza e affidabilità nei veicoli aerei e spaziali*, Atti dell'Istituto di Navigazione, Roma, 1975.

Programmi del PSN proposti e sostenuti dall'iniziativa del Prof. G. Colombo

Luciano GUERRIERO

Da ormai sette anni ho il privilegio e l'onere di dirigere il Piano Spaziale Nazionale, cui il nome di Giuseppe Colombo è indissolubilmente legato per i suoi importanti contributi che hanno portato a definire alcuni tra i più significativi progetti oggi in avanzata fase di sviluppo.

Durante i primi quattro anni di vita del Piano Spaziale Nazionale Bepi Colombo fu elemento trainante nella definizione dei programmi più innovativi e fondamentale fu la sua familiarità con le Istituzioni spaziali più prestigiose degli Stati Uniti e l'alta considerazione di cui giustamente godeva. Tutto ciò favorì lo stabilirsi di un solido ed efficace rapporto di collaborazione tra la NASA e il Piano nazionale italiano.

Proprio per il mio ruolo, ho potuto così essere testimonio privilegiato dell'azione del prof. Colombo per il programma spaziale nazionale e sono perciò grato agli organizzatori di questo importante Simposio per avermi offerto la possibilità di portare questo contributo al ricordo di questo geniale scienziato prematuramente scomparso.

Chi ha avuto la fortuna di conoscere personalmente Bepi Colombo e di lavorare con lui, sa quanto in lui la fantasia, l'entusiasmo per le meraviglie del creato, l'intuizione e la genialità fossero accompagnate da rigore scientifico, da onestà intellettuale e dall'estrema franchezza con cui respingeva ogni compromesso culturale.

Per meglio capire il legame tra il pensiero di Colombo ed alcune delle iniziative del Piano Spaziale, possiamo forse farci guidare da alcuni suoi scritti di carattere generale. Mi sono così riletto, tra le altre sue cose, un suo discorso all'Accademia Pontificia intitolato «Dal dialogo di Galileo ai dialoghi moderni» in cui si ritrova il suo interesse ed un gusto particolare per problemi scientifici fuori delle grandi mode. Particolare fascino conservava per lui il problema della dinamica dei pianeti e della stessa Terra, alla luce dei nuovi livelli di precisione offerti dalle tecnologie spaziali. Le nuove osservazioni rese disponibili dai satelliti intorno alla Terra ed alle sonde interplanetarie mostravano un nuovo livello di complessità nei fenomeni e richiedevano nuove spiegazioni.

Anche per il nostro pianeta nuove prospettive di modellizzazione de-

rivano dalle misure spaziali, sia per i fenomeni che interessano la dinamica delle masse fluide degli oceani e dell'atmosfera nella loro interazione mutua e con la radiazione solare e che determinano il clima e le sue variazioni, sia per la stessa dinamica della Terra solida che mostra una nuova complessità nelle variazioni della velocità di rotazione e nel moto del polo come pure nella lentissima dinamica della tettonica a placche.

Dall'interesse di Colombo per questa classe di fenomeni nasce il programma di geodesia spaziale del Piano Spaziale Nazionale. I suoi rapporti con la NASA, con il Jet Propulsion Lab, e con lo Smithsonian Astrophysical Observatory permettono una rapida integrazione dell'iniziativa italiana nel programma di Crustal Dynamics della NASA con l'installazione a Matera della stazione di Laserr Ranging, resa rapidamente operativa grazie all'intervento della regione Basilicata ed all'affidamento operativo alla soc. Telespazio, con risultati veramente eccellenti.

Il programma vede ormai avviate a conclusione altre iniziative che porteranno la potenzialità del nostro paese a livelli di alta competitività. Tra questa va ricordata la progettazione e la realizzazione di un sistema Laser Ranging trasportabile di nuova generazione, sotto la responsabilità della Soc. CISE e con il contributo di molte altre aziende italiane specializzate nei settori dell'ottica, della meccanica e dell'elettronica.

Sempre per la Geodesia Spaziale è in procinto di essere installato a Matera un radiotelescopio di alta precisione, dotato di un disco parabolico di 22 metri di diametro, il primo ad essere interamente progettato e costruito dall'industria nazionale sia per la parte elettronica che per quella meccanica. La responsabilità del programma è della Selenia Spazio, con il contributo della soc. SAE per la parte meccanica. La co-locazione a Matera di una stazione di Laser Ranging di nuova generazione e di un radiotelescopio per misure di geodesia con tecniche di VLBI renderanno la stazione di Matera uno dei caposaldi mondiali per misure sulla dinamica della Terra Solida. In particolare posizione di Matera al centro del mediterraneo, zona altamente sismica a causa dell'interazione locale di più placche tettoniche, garantisce a questa iniziativa una rilevanza strategica per le future campagne di misura dedicate a questa parte del globo.

A completamento delle attività di geodinamica a Matera, grazie anche alla presenza di sistemi accuratissimi per la misura del tempo (Maser ad Idrogeno) necessari per le misure radio-interferometriche di VLBI, è stata approvata dal CIPE la proposta di collocare a Matera anche una grande antenna parabolica per partecipare alle attività del Deep Space

Network per la navigazione delle sonde interplanetarie e sono stati avviati sviluppi per realizzare con l'industria nazionale ricevitori GPS/Navstar per applicazioni geodinamiche. La responsabilità in quest'ultimo caso è stata affidata alla Soc. Elettronica.

Proprio qui a Torino, negli impianti della Soc. Aeritalia stanno però nascendo per le realizzazioni più importanti del Piano nazionale per quanto riguarda la partecipazione italiana al programma mondiale di Geodinamica. Si tratta del satellite geodetico LAGEOS II che verrà messo in orbita dallo Space Shuttle mediante il stato propulsivo IRIS.

In questo programma, regolato da un MOU con la NASA, sono impegnate altre importanti industrie nazionali, quali SNIA, Microtecnica, Laben, Fiar, etc. Il satellite LAGEOS, dotato di un sistema di prismi retroriflettori per la misura accurata della distanza da terra mediante impulsi laser inviati da stazioni come quella di Matera, verrà posizionato su un'orbita circolare a circa 6000 Km di altezza, su un piano orbitale diverso da quello dell'attuale LAGEOS della NASA, in modo da aumentare l'efficienza dell'intera rete geodetica mondiale e di ottimizzare le misure sull'area mediterranea.

Satelliti del tipo LAGEOS, per la loro forma perfettamente sferica e compatta, l'elevata massa, l'elissoide di inerzia perfettamente sferico e con il centro di massa coincidente con il centro di figura, sono ideali per una accurata modellizzazione dell'orbita nel campo gravitazionale terrestre. Le deboli perturbazioni dovute alla radiazione solare, all'albedo terrestre, al frenamento da parte dell'atmosfera residua, alle correnti parassite, etc. sono valutabili con precisione. È interessante ricordare come un effetto particolare era stato riscontrato nell'orbita di LAGEOS, corrispondente ad una diminuzione della distanza di apogeo di circa un millimetro al giorno (su 12000 Km di raggio), e ciò aveva attratto l'attenzione di Colombo che si preoccupava perché tale effetto non era prevedibile sulla base dei modelli allora esistenti. Una delle ipotesi avanzata per giustificare tale rallentamento era legata ad un effetto di Relatività Generale che prevede un trascinamento del sistema inerziale causato dalla rotazione della massa terrestre. Proprio questo effetto gravito-magnetico o di Lense-Thirring è oggi all'attenzione del Piano Spaziale Nazionale che sta esaminando un aproposta per un terzo LAGEOS da immettere su un'orbita perfettamente simmetrica rispetto a quella di LAGEOS I, in modo da poter cancellare gli effetti di processione introdotti dal momento di quadrupolo del campo gravitazionale terrestre e da altre perturbazioni ed isolare così il termine relativistico, mai verificato ancora sperimentalmente.

Un altro discorso del prof. Colombo che ritengo particolarmente il-

luminante è quello pronunciato a Williamsburg, in occasione della consegna della medaglia da parte della NASA per gli straordinari contributi che lo scienziato italiano aveva dato a molti programmi dell'Ente spaziale americano. Discorso particolarmente toccante che per certi aspetti considero quasi un testamento spirituale dell'uomo che già sapeva di essere inesorabilmente condannato e che riesaminava la lunga esperienza personale ed esponeva complimenti e critiche alla grande NASA con una franchezza del tutto sconcertante, che solo a lui poteva essere concessa.

L'argomento era legato al futuro del Tethered, il satellite appeso al filo, che rappresentava l'idea che più a lungo Colombo aveva coltivato prima di vederla trasformata nel programma più prestigioso della collaborazione tra la NASA ed il programma spaziale italiano. In quella occasione Colombo spiegò la difficoltà incontrata nel far capire un concetto particolarmente semplice con l'abitudine a considerare lo Spazio come una estensione dell'ambiente terrestre dominato da fenomeni violenti ed aleatori, mentre occorre invece tener conto che i sistemi nello spazio vanno visti come regolati dalle leggi della meccanica celeste, dove le leggi fondamentali sono inerzia e gravitazione, con forze deboli e perfettamente deterministiche e pertanto facilmente modellizzabili.

Colombo, per la sua lunga esperienza scientifica di meccanico celeste, aveva dedicato particolare attenzione alla possibilità di realizzare sistemi spaziali in grado di utilizzare appieno le proprietà dinamiche di questo nuovo ambiente. Particolarmente interessanti erano per lui le proprietà di stabilità intorno ai punti lagrangiani, gli effetti stabilizzanti del gradiente di gravità sui sistemi estesi, la spinta del vento solare e della pressione di radiazione, gli scambi di energia e momento ottenibili utilizzando risonanze orbitali con le armoniche elevate del campo gravitazionale ed infine le azioni frenanti dei bordi dell'atmosfera.

Le grandi strutture spaziali che Colombo vedeva nel futuro delle attività spaziali, e per le quali riteneva indispensabile il binomio Space Shuttle - Stazione Spaziale permanente, faranno pieno uso di queste proprietà dinamiche ed in questo senso il sistema Tethered, caratterizzato da due masse orbitanti collegate da un lungo filo tenuto in tensione dal gradiente di gravità rappresenta la prima realizzazione di una struttura spaziale dalle dimensioni di diverse decine di chilometri.

Il primo esperimento con un sistema Tethered è attualmente in corso di realizzazione come collaborazione tra il Piano Spaziale Nazionale e la NASA. Il satellite sferico del diametro di 1.6 metri e di circa 500 chili di massa viene realizzato sotto la responsabilità della Soc. Aeritalia e con il contributo delle altre maggiori aziende aerospaziali italiane. Il filo conduttore che collegherà il satellite allo Space Shuttle durante il volo

orbitale si estenderà per venti chilometri e movendosi nel campo magnetico terrestre diventerà sede di forza elettromotrice indotta permettendo così di chiudere correnti elettriche nel plasma della magnetosfera lungo le linee di forza del campo magnetico terrestre.

Sotto responsabilità italiana sono pure le apparecchiature del cosiddetto Core Equipment per il controllo delle correnti e delle f.e.m. indotte come pure della dinamica del satellite. A bordo del satellite e su dei booms estendibili saranno integrati alcuni esperimenti elettrodinamici che permetteranno la misura diretta della carica spaziale nella regione intorno al satellite, la densità di elettroni, la densità e la temperatura del plasma, la densità e la distribuzione di energia degli ioni nonché l'intensità del campo elettrico e di quello magnetico.

Il primo esperimento con il sistema Tethered costituirà certamente una tappa fondamentale nella storia delle attività spaziali. Altri voli ed altri esperimenti seguiranno. Un secondo volo è già stato concordato con la NASA per far scendere un satellite con un filo lungo ben cento chilometri ed esplorare in tal modo gli strati più alti dell'atmosfera, estremamente importanti per molti degli aspetti della dinamica, della termodinamica e della chimica dell'atmosfera e difficilmente studiabili con i metodi tradizionali. Dal secondo Tethered sarà anche possibile ottenere importanti informazioni sulla fluidodinamica in condizioni tipiche del rientro nell'atmosfera di satelliti e sonde, condizioni non realizzabili in laboratorio.

Sotto la spinta di Colombo, la NASA ed il Piano Spaziale italiano hanno avviato anche una serie di studi per valutare le potenzialità dei sistemi a filo per varie applicazioni spaziali, anche in vista della futura Stazione Spaziale permanente e delle nuove missioni sia scientifiche che applicative che questa renderà possibili.

Queste applicazioni del Tethered riguardano problemi di Rendez-vous e Docking, problemi di dinamica orbitale per trasferire a carichi utili energia e momento angolare, problemi di stabilizzazione dinamica di grandi sistemi, problemi di gravità artificiale a livello controllabile, etc.

In vari simposi internazionali che si sono alternati negli Stati Uniti ed in Italia, una moltitudine di idee interessanti è stata portata all'attenzione di una comunità di specialisti sempre più vasta ed entusiasta.

Non sembra esserci dubbio sull'importanza che questa idea di Colombo avrà per lo sviluppo delle grandi strutture spaziali del futuro. Basta esaminare il recente rapporto Paine, predisposto da un gruppo di esperti qualificati per il Congresso americano, per constatare quanta strada abbia fatto ormai questo nuovo modo di pensare, certamente dovuto agli insegnamenti che Colombo ci ha lasciato come sua principale eredità.

In questa mia breve relazione mi sono limitato a toccare solo gli aspetti più rilevanti del ruolo che il prof. Colombo ha svolto nell'impostazione e nello sviluppo del Piano Spaziale Nazionale italiano. Certamente tutti i programmi del Piano, anche quelli che non ho menzionato, hanno risentito positivamente della sua esperienza, della sua critica e della sua capacità di cercare sempre le soluzioni che più assecondano le tendenze spontanee della natura.

La prematura scomparsa del prof. Colombo, maestro ed amico di molti di noi, ha creato certamente un vuoto difficilmente colmabile nella comunità spaziale italiana e mondiale. Il Piano Spaziale Nazionale ha avuto il privilegio di farsi guidare da lui in alcune scelte che risultano oggi quanto mai significative. Il portare a compimento questi programmi rappresenta oggi non soltanto un dovere per i responsabili della attività spaziale italiana ma anche il miglior modo per onorare la memoria di questo eccezionale scienziato.

Geophysical parameter determination using satellite ranging

Byron D. TAPLEY (*)

Abstract. *Modern geophysical parameter determination rests on the conceptual foundations developed by Gauss to compute the orbits of planetary bodies. The technique has been essential in the determination of orbits and trajectories of near-earth and interplanetary satellites. As measurement precision and computing capability have increased, applications of the technique to earth-orbiting satellites have evolved from the orbit determination to include areas of geodesy, geophysics and oceanography. The following discussion summarizes the approach, describes some of the recent applications, and discusses some of the recent results.*

Introduction

During the past two decades, highly accurate ranging to earth-orbiting satellites has been used to determine their orbits. The measurements include satellite laser range measurements [Plotkin et al., 1973] and satellite altimeter height measurements [Tapley et al., 1982], which are determined by time-of-flight measurements, biased range measurements, which are determined by measurements of carrier-phase [Remondi, 1984] or by interferometric determination of time delay [Counselman and Steinbricher, 1981]. Although there has been a continuing improvement in the precision of the range measurements with current ranging precision reported at the 1-cm level, the accuracy of the orbits has been limited by error in the dynamic model used to describe the satellite's motion and the measurement model used to interpret the range measurements. To improve the accuracy of the orbits, the ranging data have been used to improve the models for the forces which influence the satellite's motion. The most important forces include those due to the inhomogeneous mass distribution of the central body, time variation in the central body mass distribution, and surface forces due to atmospheric drag and radiation pressure. The time variations in the locations of the observation

(*) Center for Space Research, The University of Texas at Austin, Austin, Texas 78712 USA.

points and the effects of atmospheric refraction are other important error sources. Accurate models for each of these effects are required if the orbit accuracy is to be commensurate with current measurement precision. In the process of improving the factors which limit the ability to compute the orbits of near-earth satellites, significant advances have been made in our knowledge of the earth's solid-body geopotential and the influence of solid body and ocean tides on the motion of near-earth satellites [Marsh et al., 1988; Tapley et al., 1988]. The requirement for accurate determination of the coordinates of the topocentric tracking stations has dictated an associated requirement for substantial improvement in the knowledge of the global station positions and the plate tectonic and earth rotation models which influence the time variation in the station locations [Tapley et al., 1985]. The use of dynamically continuous multi-year arcs has allowed the study of the effects of secular changes in the earth's mass distribution due to post-glacial rebound [Yoder et al., 1983] and the effects of relativity [Ries et al., 1988]. The approach to using high precision range measurements to near-earth satellites to perform these studies, along with some of the contributions from these investigations, are described in the subsequent discussion.

Orbit Determination Theory

The techniques used to determine the orbits of near-earth satellites reside in the pioneering work of Gauss [1809] who developed the original theory of least squares parameter estimation to determine orbits of the planets in the solar system and the trajectory of comets which pass through the solar system. The differential correction and least squares concepts proposed by Gauss provide the conceptual basis on which modern orbit determination procedures are based. These ideas have been placed on a firm theoretical and computational foundation by advances in probability theory and statistics, numerical approzimation theory, linear systems theory, and the development of the digital computer [Fisher, 1912; Wiener, 1942; Kolmogorov, 1941; Maybeck, 1974]. For current applications, the solution to the orbit determination problem involves four fundamental elements: (1) a set of differential equations which describe the motion of the satellite, (2) a numerical integration procedure to obtain a solution to the differential equations, (3) accurate observations of the satellite's motion, and (4) an appropriate estimation method which combines the results of the first three to yield an estimate of the satellite's position, velocity and appropriate model parameters (e.g.,

the satellite state) [Tapley, 1973; Sorenson, 1970].

For most applications, the equations of motion for the satellite are based on Newton's Second Law. For some applications, corrections to account for the effects of General Relativity are required [Ries et al., 1988]. If the satellite's state is known at some epoch, t_o , the equations of motion can be integrated to obtain a prediction of the satellite's motion. The observations, or tracking data, can be used to either verify the accuracy of the prediction or to correct it. The tracking data are scalar measurements which depend on the satellite state and the location of the observer. If the tracking station is located on the earth's surface, models for the orientation of the earth's angular velocity in space (precession and nutation) and the orientation of earth's angular velocity relative to the body-fixed coordinate system (earth orientation parameters, polar motion and UT1) are required to describe the tracking station motion in the inertial reference frame. Finally, the time rate of change of the locations of the tracking stations on the surface of the earth require that a model for the earth's tectonic plate motion be defined.

Errors in either the dynamic model or the measurement model will lead to a discrepancy between the calculated value of the observation, based on the predicted motion, and the actual observation.

This discrepancy, observed over a sufficient time interval, can be used to improve the knowledge of the parameters in the dynamic and measurement models, as well as the accuracy of the trajectory. The techniques used to accomplish this objective fall in the general category of nonlinear parameter estimation. The essential ideas in this approach are outlined in the following discussion.

The Equations of Motion

The differential equations which govern the motion of a near-earth satellite corrected to include the effects of relativity can be stated as follow:

$$\ddot{\bar{r}} = -\frac{\mu \bar{r}}{r^3} + \bar{R} + T[\bar{G} + \bar{D} + \bar{E}] + \bar{S} + \bar{P} + \bar{m} \tag{1}$$

where μ is the gravitationa constant for the earth

$\bar{R}$ is the effect of relativity

$\bar{G}$ is the effects due to all non-spherical gravitational forces due to the earth's nonuniform and time-varving mass distribution

$\bar{D}$	is the effect of atmospheric drag
$\bar{E}$	is the effect of earth reflected and radiated energy
$\bar{S}$	is the effect of solar radiation
$\bar{P}$	is the effect of the lunar and planetary perturbations
$\bar{m}$	is the effects of any unmodeled forces which act on the satellite
T	is the 3×3 transformation matrix which relates geocentric body-fixed coordinates to inertial coordinates.

Explicit expressions for the components of acceleration in Eq. (1) are given in Appendix A. For numerical integration purposes, $\bar{r}$ and its second derivative are expressed in a nonrotating, earth-centered coordinate system, e.g., the axis of the geocentric coordinates are parallel to the axis of an inertial reference frame.

The general form for Eq. (17 can be expressed as

$$\ddot{\bar{r}} = \bar{f}(\bar{r}, \dot{\bar{r}}, \bar{p}, t) + \bar{m}(t) \tag{2}$$

where $\bar{r}$ is the geocentric position of the satellite, $\dot{\bar{r}}$ is the geocentric velocity, t is ephemeris time, which is assumed to be identical to international atomic time TAI, and $\bar{p}$ is a set of constant parameters which appear in the mathematical model for the equations of motion. The vector, $\bar{m}(t)$, represents the unmodeled accelerations which act on the spacecraft due to either a functionally incorrect description of the various forces which affect the spacecraft or inaccurate values for the constant parameters, $\bar{p}$, which appear in the force model.

Information related to the structure and time variation of the earth's compositioin can be inferred by observing the behaviour of the satellite's motion over long time intervals. For example, the observation of the motion of a satellite can be obtained through laser range measurements which can be modeled as

$$Y_i = G(\bar{r}_i, \bar{r}^j_{s_i}, t_i)_i + \epsilon_i \tag{3}$$

For a satellite laser range measurement, Y_i is the measured time interval between transmission and reception of a laser pulse at a tracking station, and $\bar{r}^j_{s_i}$ is the coordinates of the j the tracking station. As shown in Figure 1, during the measured time interval, the pulse travels to the satellite, where it is reflected by corner cube reflectors on the satellite and returns to the laser tracking station. Consequently, one-half of the round-trip time interval, multiplied by the speed of light, will pro-

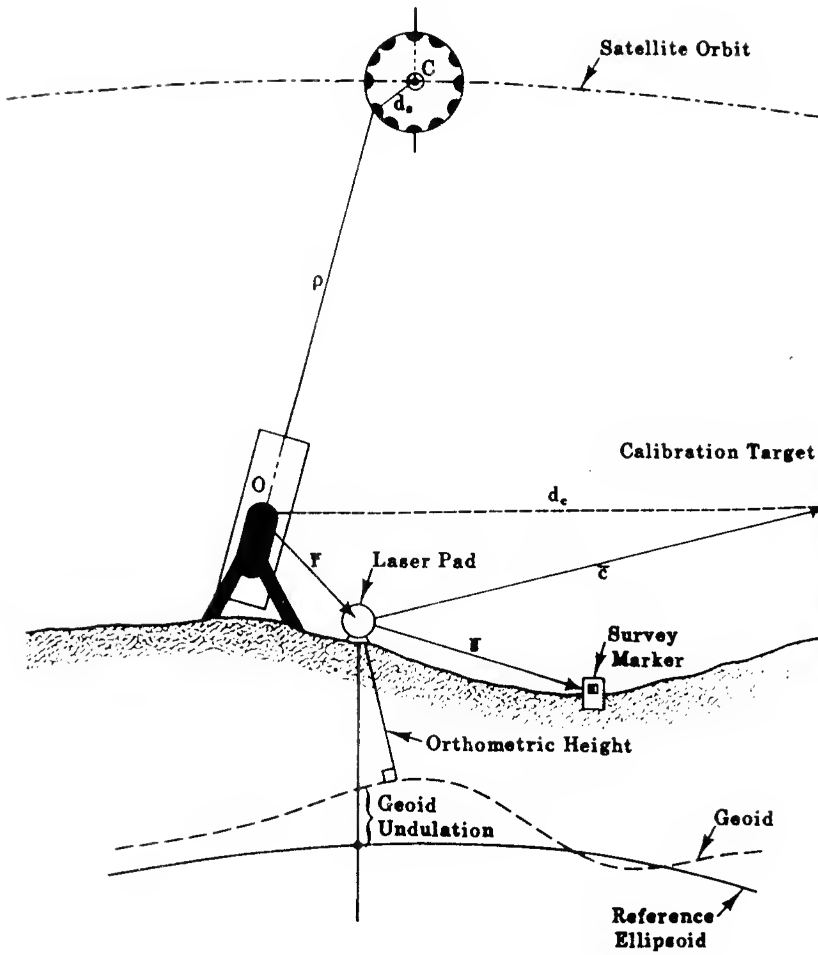


Figure 1

duce a measurement of the distance between the satellite and the tracking station at the time the signal arrives at the satellite. This value, when corrected for the effects of atmospheric refraction and other instrument and measurement delays, is Y_i in Eq. (1). Similar relations hold for altimeter height measurements and doppler or phase inferred range difference measurements. The nonlinear function $G(\bar{r}_i, \bar{r}^j_{s_i}, t_i)$ is the scalar distance between the jth tracking station, $\bar{r}^j_{s_i}$ and the satellite at the observation epoch, t_i. That is, for the laser range measurement

(4) $$G(\bar{r}_i, \bar{r}^j_{s_i}, t_i) = \| \bar{r}(t_i) - \bar{r}^j_s(t_i) \|$$

As noted in Figure 1, the laser tracking station positions are usually expressed in earth-fixed coordinates. The relation between earth-fixed and quasi-inertial geocentric nonrotating coordinates can be expressed as

$$(5) \qquad \bar{r}_s^j = B\ N\ \bar{R}_s^j$$

where $\bar{R}_s^j$ is the position of the ith station in the body-fixed system, $\bar{r}_s^j$ is the position in the quasi-inertial nonrotating system, the 3×3 matrix B contains the effects of earth rotation, e.g., the polar motion, x_p, y_p and UT1 variations, and the 3×3 matrix N contains the effects of precession and nutation of the earth's angular monumentum vector relative to inertial space.

The tectonic deformation information as well as the signatures used to extract polar motion and earth rotation will be present in Eqs. (4) and (5). To account for these effects, the motion of the tracking stations in the body-fixed or terrestrial reference frame can be modeled as

$$(6) \qquad \bar{R}_s^j(t) = \bar{R}_{s_o}^j + \bar{\Omega}^k \times \bar{R}_{s_o}^j (t - t_o) + \Delta R_s^j(t)$$

where t_o is a reference epoch
$\bar{R}_{s_o}^j$ is the body-fixed coordinates of the jth station at t_o
$\bar{\Omega}^k$ is the angular velocity of the ith tectonic plate expressed in the terrestrial reference frame
$\bar{R}_s^j(t)$ is the coordinates of the jth station at the time t
$\Delta R_s^j(t)$ is the change in the jth station coordinates due to local deformation, subsidence, upflit, etc.

With a sufficient distribution of the tracking stations on the k th plate, Eq. (6) can be used to estimate the plate rotation parameters, $\bar{\Omega}^k$ along with the epoch station coordinates, $\bar{r}_{s_o}^j$. To achieve an accurate estimate of $\bar{\Omega}^k$, at least three stations should be located on the kth tectonic plate.

Eqs. (2) and (3) form a system of nonlinear equations which describe the time evolution of the satellite's motion and the relation of the observations to the motion. To complete the formulation, a procedure for using a sequence of observations, Y_i; $i = 1, \ldots, n$, to estimate the state of the satellite at some time epoch, t_o, must be determined. The state consists of the position, $\bar{r}_o$, the velocity, $\dot{\bar{r}}_o$, and the measurement and dynamic model parameters which are to be estimated. The estimate of the measurement and dynamic model parameters provide the coupling of the satellite orbit determination problem to the field of geodesy, sin-

ce the model parameters to be estimated will include the position of the tracking stations and the earth's geopotential and tide model parameters. Studies of this nature fall in the general area of Satellite Geodesy.

The usual technique for estimating the parameters of interest is based on the application of linear estimation techniques to the nonlinear estimation problem outlined above. The approximations to enable this transition are outlined in the following discussion.

Linearization of the Orbit Determination Process

If a reasonable reference trajectory is available and if X, the true trajectory, and X^*, the reference trajectory, remain sufficiently close throughout the time interval of interest, the trajectory for the actual motion can be expanded in a Taylor's series about the reference trajectory at each point in time to obtain a set of linear differential equations with time dependent coefficients. Using a similar procedure to expand the nonlinear observation state relation, a linear relation between the observation deviation and the state deviation can be obtained. Then, the nonlinear orbit determination problem can be replaced by a linear orbit determination problem in which the deviation from some reference trajectory is to be determined [Tapley, 1973].

Let

$$x(t)=X(t)-X^*(t) \qquad y(t)=Y(t)-Y^*(t) \tag{7}$$

where $X^*(t)$ is a specified reference trajectory and $Y^*(t)$ is the value of the observation calculated by using $X^*(t)$. Then, substituting Eq. (7) into Eqs. (5) and (6), expanding in a Taylor's series, and neglecting terms of order higher than the first leads to the relations

$$\begin{aligned} \dot{x} &= A(t)\,x \\ y_i &= \tilde{H}_i x_i + \epsilon_i \quad (i=1, \dots, l) \end{aligned} \tag{8}$$

where

$$A(t)=\frac{\partial F}{\partial X}(X^*, t) \qquad \tilde{H}=\frac{\partial G}{\partial X}(X^*, t) \tag{9}$$

The general solution to the first of Eqs. (8) can be expressed as

$$(10) \qquad x(t) = \Phi(t, t_k)\, x_k$$

where $x(t)$ is the value of x at a specific time t, and where the $n \times n$ state transition matrix $\Phi(t, t_k)$ satisfies the differential equation:

$$(11) \qquad \dot{\Phi}(t, t_k) = A(t)\,\Phi(t, t_k), \quad \Phi(t_k, t_k) = I.$$

Using Eq. (9), the second of Eqs. (7) may be written in terms of the state at t_o as

$$(12) \qquad y_i = \tilde{H}_i\,\Phi(t_i, t_o)\,x_o + \epsilon_i, \quad i = 1, \ldots, m.$$

If the following definitions are used

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \qquad H = \begin{bmatrix} \tilde{H}_1\Phi(t_1, t_o \\ \vdots \\ \tilde{H}_m\Phi(t_l, t_o) \end{bmatrix} \qquad \epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_m \end{bmatrix}$$

and if the subscript on x_o is dropped, then Eqs. (12) can be expressed as follows:

$$(14) \qquad y = Hx + \epsilon$$

where, if m is the total number of observations, y is an $m \times 1$ vector, x is an $n \times 1$ vector, ϵ is an $m \times 1$ vector, and H is an $m \times n$ mapping matrix. For the cases of interest, the essential condition $m > n$ is satisfied. It is usually assumed that the observations error vector, ϵ, satisfies the *a priori* statistics, $E[\epsilon] = 0$ and $E[\epsilon\epsilon^T] = W^{-1}$. By scaling each term in Eq. (14) by $W^{1/2}$, the condition

$$(15) \qquad W^{1/2} E[\epsilon\epsilon^T]\, W^{T/2} = W^{1/2}\, W^{-1}\, W^{T/2} = I$$

is obtained.

Least Squares Solution via Orthogonal Transformation

An approach to obtain the best estimate of x, given the linear observation-state relations (Eq. (14)), which avoids some of the numerical problems encountered in the conventional normal equation approach is described in the following discussions. The method obtains the solution by applying successive orthogonal transformations to the linear equations given in Eq. (14). Consider the quadratic performance index

$$J = 1/2\ \| W^{1/2}(Hx - y) \|^2 = 1/2 (Hx - y)^T W (Hx - y). \tag{16}$$

The solution to the weighted least squares estimation problem (which is equivalent to the minimum variance and the maximum likelihood estimation problem, under certain restrictions) is obtained by finding the value $\hat{x}$ which minimizes Eq. (16). Without changing the value of J, Eq. (16) can be expressed as

$$J = 1/2\ \| Q W^{1/2}(Hx - y) \|^2 \tag{17}$$

where Q is an arbitrary $m \times m$ orthogonal matrix. Now, if Q is selected such that

$$Q W^{1/2} H = \begin{bmatrix} R \\ O \end{bmatrix} \quad Q W^{1/2} y = \begin{bmatrix} b \\ e \end{bmatrix} \tag{18}$$

where R is $n \times n$ upper-triangular, O is an $(m-n) \times n$ null matrix, b is an $n \times 1$ vector, and e is an $(m-n) \times 1$ vector, then Eq. (17) can be written as

$$J(x) = 1/2\ \| Rx - b \|^2 + 1/2\ \| x \|^2. \tag{19}$$

The value of x which minimizes (16) is obtained by the solution

$$R\hat{x} = b \tag{20}$$

and the minimum value of the performance index becomes

$$J(\hat{x}) = 1/2\ \| e \|^2 = 1/2\ \| y - H\hat{x} \|^2. \tag{21}$$

That is, e provides an estimate of the residual error vector.

The procedure is direct and for mechanization requires only that a convenient computation procedure for computing $QW^{1/2}H$ and $QW^{1/2}y$ be available. The two most frequently applied methods are the Givens method, based on a sequence of orthogonal rotations, and the Householder, method, based on a series of orthogonal reflections [*Lawson and Hanson*, 1963]. In addition to the expression for computing the estimate, the statistical properties of the error in the estimate, R, are required. If the error in the estimate, η , is defined as

$$\eta = \hat{x} - x \tag{22}$$

it follows that

$$E[\eta] = E[\hat{x} - x] = E[R^{-1}b - x]. \tag{23}$$

Since

$$QW^{1/2}y = QW^{1/2}Hx + QW^{1/2}\epsilon$$

leads to

$$b = Rx + \tilde{\epsilon} \tag{24}$$

it follows that

$$E[\eta] = E[R^{-1}(Rx + \tilde{\epsilon}) - x] = E[R^{-1}\tilde{\epsilon}] \tag{25}$$

As noted in Eq. (15), if the observation error, ϵ, is unbiased, $\tilde{\epsilon} = QW^{1/2}\epsilon$ will be unbiased and

$$E[\eta] = 0. \tag{26}$$

Hence, $\hat{x}$ will be an unbiased estimate of x. Similarly, the covariance matrix for the error in x can be expressed as

$$\begin{aligned} P = E[\eta\eta^T] &= E[R^{-1}(Rx + \tilde{\epsilon} - x)(Rx + \tilde{\epsilon} - x)^T R^{-T}] \\ &= E[R^{-1}\tilde{\epsilon}\tilde{\epsilon} R^{-T}] = R^{-1}E[\tilde{\epsilon}\tilde{\epsilon}^T]R^{-T}. \end{aligned} \tag{27}$$

If the observation error, ϵ, has a statistical covariance defined as $R[\epsilon\epsilon^T] = W^{-1}$, the estimation error covariance matrix is given by

$E[\tilde{\epsilon}\,\tilde{\epsilon}^{\,T}] = W^{1/2}E[\epsilon\epsilon^T]\,W^{T/2} = W^{1/2}\,W^{-1}\,W^{T/2} = I$. Consequently, relation (27) leads to

$$(28) \qquad P = R^{-1}R^{-T}.$$

It follows then that the estimate of the state and the associated state error covariance matrix are given by the expressions

$$(29) \qquad \hat{x} = R^{-1}b, \quad P = R^{-1}R^{-T}$$

where x represents the error in the initial condition, $\bar{r}_o$ and $\dot{\bar{r}}_o$, the error in the dynamic model parameters and the error in the measurement model parameters.

Applications to Satellite Geodesy

The field of geodesy is concerned with determining the geometry and gravity field of the earth, the location of points on the earth's surface, and the determination of the time variations in these quantities. The application of satellites to the field of geodesy was one of the earliest uses of space technology. The use of both ranging to satellites from points on the earth's surface as well as ranging to the earth's surface from radar altimeters carried onboard satellites has provided a comprehensive view of the earth, which is not achievable by conventional ground-based methods. Although early contributions were obtained using doppler frequency measurements to the Navy navigation satellites, satellite laser ranging (SLR) and satellite altimetry have been the primary systems used for these studies since 1976. The laser geodynamics ranging satellite, Lageos, was launched in May 1976 [Johnson et al., 1976], and this event, coupled with the development and deployment of a global tracking network of satellite laser ranging systems in 1978, led to significant contributions to the global geodesy. The SLR network deployment was stimulated by the launch of Seasat, which specified satellite laser ranging as the tracking technique to satisfy a 10-cm radial orbit accuracy requirement. The altimeter carried on Seasat had a ranging precision of 5 cm and gave the first global view of the ocean topography [Tapley et al., 1982]. Although the best orbit accuracy for Seasat, computed more than a decade after launch, is around 30 cm rms, the detailed information on the marine geoid provided by the Seasat altimeter is still being used to discover new features of the time-varying ocean circulation and

variability as well as the marine geoid. Furthermore, the laser network developed to support the Seasat mission provided the basis for the collaboration required to establish a global network of satelite laser ranging stations required to support the international Crustal Dynamics Project [Coates et al., 1985] and important regional campaigns, including the Wegener, SAFE and Western North American campaigns. The objectives of the Crustal Dynamics Project are to study the motions of the earth's tectonic plates and the overall dynamic motion of the earth with the specific objective of improving the knowledge of

1. Regional deformation and strain accumulation related to large earthquakes in the plate boundary regions, such as the western North America region.
2. The contemporary motions of the North American, Pacific, South American, Nazca, European and Australian Plates.
3. The internal deformation of continental and oceanic lithospheric plates, with particular emphasis on North America and the Pacific.
4. The rotational dynamics of the earth and their possible correlation to earthquake, plate motions and other geophysical phenomena.

The international SLR network played a vital role in the MERIT Campaign and established satellite laser ranging as one of the two methods of choice for determining the earth's rotation [Wilkins, 1980; Tapley, 1983; Robertson et al., 1985]. The other technique is based on very long baseline interferometry (VLBI).

The SLR system has provided accurate measurements for orienting the earth within a quasi-inertial reference system since 1980. Global data taken over 3-day intervals provide a suitable means for solving directly for the earth's polar motion and UT1 [Schutz et al., 1984]. Current accuriacies, which are routinely achieved for 3-day mean values, are 2 milliarcsec for the x and y coordinates of the rotations axis and 0.1 msec for change in length of day [Robertson et al., 1985; Schutz et al., 1985; Tapley et al., 1985]. Figure 2 shows the earth rotation solutions for SLR during the period from 1980 through 1984. The black diamonds are independent determinations from the VLBI systems which did not begin regular reporting until the beginning of the MERIT Campaign in September 1983. The analysis of the long period changes in the node have been interpreted in terms of the secular and seasonal changes in the earth's gravitation field coefficient, J_2, the term which describes the earth's oblateness. The secular change in this value has been related to the viscoelastic rebound of North America, Northwestern Europe and Antarctica following the melting of large continental ice sheets subsequent to the last glacial maximum [Yoder et al., 1983; Peltier and Wu, 1983; Ru-

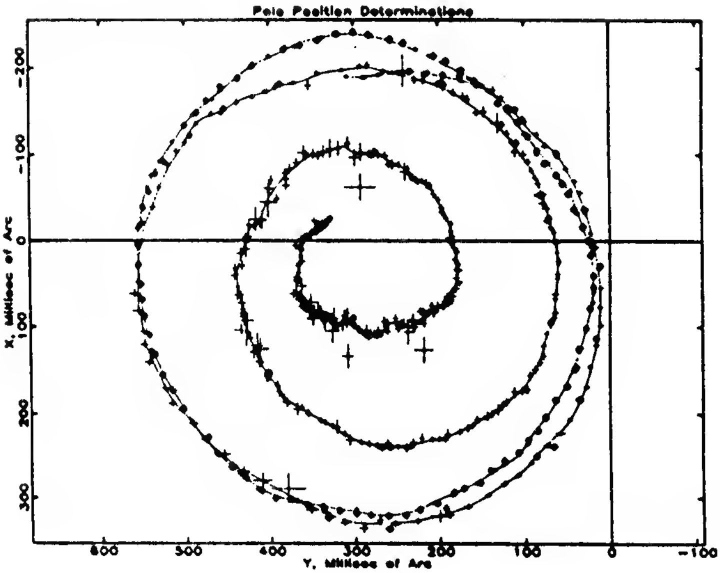


Figure 2 - SLR and VLBI polar motion solutions for period from September 1980 through December 1983 (+SLR, ♦ VLBI).

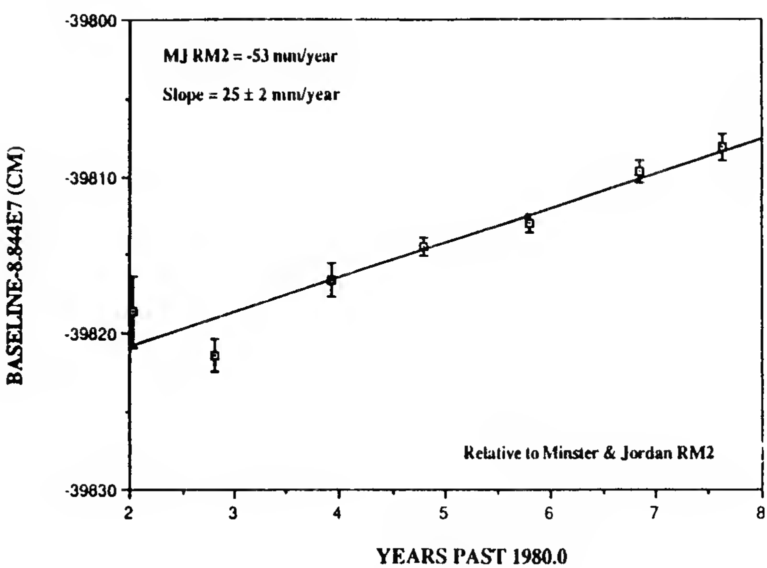


Figure 3 - Monument peak quincy annual LLA8801.

bincam, 1984; Wu and Peltier, 1984].

The accurate determination of the earth's rotation has allowed studies of the angular monumentum exchange between the solid earth and the atmosphere [Chao, 1984; Salstein and Rosen, 1985]. In addition, operating under the stimulus of the Crustal Dynamics Project, the techniques of SLR and VLBI provided the first definitive measurements to confirm the plate tectonic theory [Tapley et al., 1985; Smith et al., 1985; Clark et al., 1987]. Figure 3 shows the secular increase in the distance between two laser ranging stations located on opposite sides of the San Andreas Fault in California, U.S.A.. A six-year period of SLR measurements to Lageos were used to infer a rate of change of -25 mm/year relative to the rate of -53 mm/year predicted by geophysical measurements. The uncertainty in the slope is 2 mm/year. The results in Figure 3 indicate the degree to which both measurements and orbit computation methodology have matured during the past decade.

The Topex Mission

The need for more accurate orbits to support the recent satellite altimeter missions, Geosat, European Remote Sensing (ERS-1) satellite and the NASA-CNES Topex/Poseidon satellite [Stewart et al., 1986], has been a major stimulus for the improvement and standardization of the software systems used to compute orbits and the development of improved models for the forces which act on the satellites. The orbit accuracy requirements for the Topex/Poseidon mission are so demanding that they have been the primary stimulus for improving orbit determination methodology during the past decade. As a consequence, attention is directed to the Topex/Poseidon mission in the subsequent discussion.

With an independent determination of the satellite height from the orbit determination system and an accurate marine geoid, the altimeter height measurements can be used to infer the ocean surface topography. As a follow-on to the Seasat mission, NASA and CNES have agreed to collaborate in a joint mission to use a satellite altimeter to measure the ocean surface topography over entire ocean basins for a period of three to five years. The mission, referred to as the Topex/Poseidon mission, will use a satellite altimeter which has the ability to measure the distance from the satellite to the ocean surface with an instrument precision of approximately two centimeters. The oceanographic phenomena which influence the surface topography include currents, mesoscale

eddies, tides, storm surges and the marine geoid [Apel, 1980]. The measurements obtained by the altimeter during this mission will be integrated with subsurface oceanographic measurements and models of the ocean's density field to determine the general circulation of the ocean and its variability. The geoid undulations over the ocean are as large as 150 meters, while the maximum value of the ocean quasi-static surface topography is on the order of 1 meter. Consequently, the measurement presents a formidable challenge to the orbit determination techniques. Further rational for the mission and its role in the general ocean topography experiment are given by the *Topex Science Working Group* [1981] and Stewart et al. [1986].

For satellite altimetry applications, the satellite serves as a stable platform from which the altimeter measures the average distance from the antenna feed point to the instantaneous electromagnetic mean sea level. Figure 4 illustrates the satellite altimeter measurement. As the satellite flies over the ocean surface, the distance from the satellite to the ocean surface is inferred from the time of flight of radar pulses transmitted from the satellite and reflected from the ocean surface. The time delay

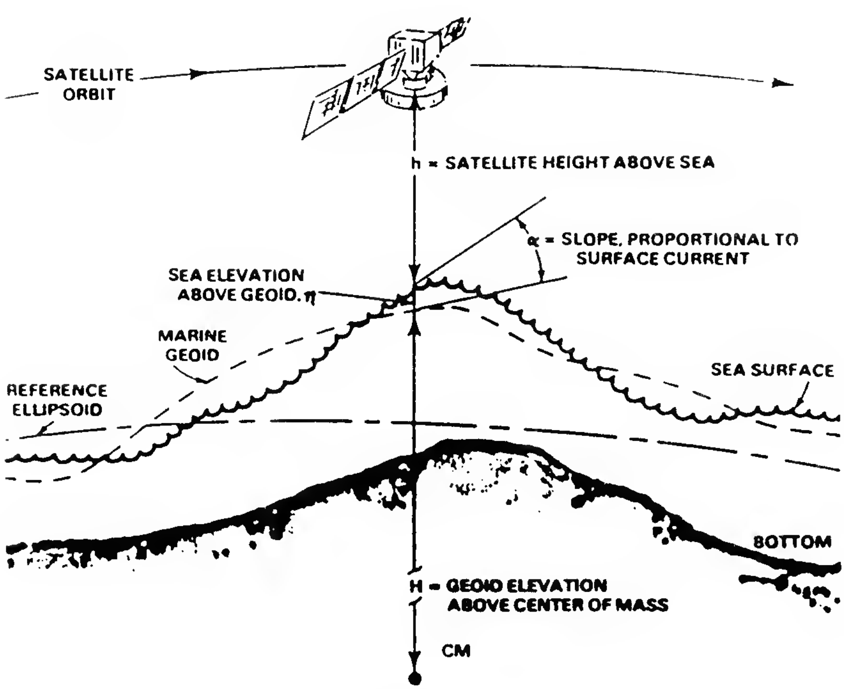


Figure 4 - The Satellite Altimeter Measurement.

is corrected for the effects of troposphere and ionosphere refraction, instrument delay, radiometric sea surface interaction and radar antenna-to-mass center offset [Tapley et al., 1982]. Given the geoid and the altimeter height measurement, the time-varying ocean surface changes can be determined.

A necessary condition to be able to use the satellite altimeter measurements in modeling the ocean surface height is that the satellite orbit be computed with an accuracy comparable to the accuracy of the satellite altimeter measurement. The science requirement for the Topex/Poseidon mission is that the radial component of the satellite's orbit must be known to the order of 13 cm rms for the mission's duration [Stewart et al., 1986], a unique and extremely difficult challenge for the precision orbit determination field. An additional requirement is that the non-averaging geographically correlated errors over ocean basins must be less than 5 cm. This imposes a very strenous orbit determination requirement and has led to a significant stimulus to improve the precision orbit determination methodology described above.

To separate variation in the satellite height from variations in the ocean surface height and to orient the ocean surface height measurements with respect to a common center-of-mass coordinate system, an independent determination of the radial component of the satellite's orbit is determined from ground-based tracking of the spacecraft. To achieve the primary measurement of the broadscale ocean circulation, the orbit must be computed with the requisite accuracy in a body-fixed coordinate system whose origin is located at the earth's mass center. This requires, not only an accurate orbit determination procedure, but a procedure for defining the terrestrial or body-fixed reference system and the location of the tracking stations in this reference system.

The orbit determination accuracy will depend on the knowledge of the dynamic forces which act on the satellite, the accuracy of the tracking station coordinates and the accuracy and frequency with which the satellite's motion can be observed. As noted in Appendix A, the dominant forces which influence the satellite's orbit are due to the gravitational attraction of the earth, sun and moon, the effects of atmospheric resistance and the influence of the direct and reflected solar radiation pressure. Each of these effects must be modeled with sufficient accuracy to meet the orbit computation requirements. Other effects which influence the orbit accuracy include geographic and temporal distribution of the tracking data and the sophistication of the software system used to determine the estimates of the orbit.

The single largest error source in the error budget for the Topex sa-

tellite is due to the error in the earth's gravity field model. In addition to contributing to the long period and secular growth in the radial orbit error, the short period error (on the order of one revolution or less) has a geographically correlated component that is particularly insidious for satellite altimetry applications [Tapley and Rosborough, 1985; Rosborough, 1986; Rosborough and Tapley, 1987]. These investigations demonstrate that the gravity field induced short period orbit error will have a component which is geographically correlated and, furthermore, it will vary in phase with the long wavelength components of the geoid error. Accurate satellite tracking using ground-based tracking systems can be used to remove the long period and resonant orbit error components. These components can be removed by breaking the ephemeris up into a sequence of independent arcs of a few days (normally two to ten days) and adjusting the satellite position and velocity at each arc epoch. The daily and short period effects cannot be eliminated in this manner, unless nearly continuous tracking of the satellite is available to allow solutions of the satellite ephemeris for arc lengths less than one orbit revolution. The only way to remove the geographically correlated orbit error is to improve the geopotential model.

Since the geopotential model error is expected to be the dominant contributor to the satellite orbit error, particular emphasis has been devoted to this error source in the pre-mission planning. Preliminary results from a multi-satellite solution for the earth's gravity field model have shown a significant improvement in the accuracy of the current model for the earth's gravity field [Marsh et al., 1988; Tapley et al., 1988]. Table 1 indicates the satellite data used in the solution reported by Tapley et al. [1988]. Table 2 shows the rms of fit to the tracking data for the new model compared with similar results from the previous best-fitting specially tailored gravity model. The gravity models developed for Topex yield substantial improvements in the orbit accuracies achieved for all satellites. For the first time, a single satellite determined gravity model will fit the data from over 25 satellites better than previously developed specially tailored models.

Future Trends

Future concerns for using space techniques will lead to increasing stringency in the orbit accuracy requirements. The global context obtained by space techniques dictate that they play a central role in monitoring long-term global change, whether it be due to geophysical, biological

Table 1. Satellite Data for the University of Texas Gravity Model, TEG-1

Satellite	Launch Date	Data	Inclination	Eccentricity	Altitude (km)
Geos-1	1965	Laser	59°	0.072	1600
BE-C	1966	Laser	41°	0.026	1130
DI-C	1967	Laser	40°	0.053	1000
DI-D	1967	Laser	39°	0.085	1200
Oscar-14	1967	Doppler	89°	0.005	1100
Geos-2	1968	Laser	106°	0.033	1400
Peole	1971	Laser	15°	0.015	650
Geos-3	1975	Laser	115°	0.002	830
Starlette	1975	Laser	50°	0.020	900
Lageos	1976	Laser	110°	0.004	5900
Seasat	1978	Laser, Doppler, Altimeter and Crossover	108°	0.002	800
Nova-1	1980	Doppler	90°	0.002	1200
Geosat	1985	Doppler, Altimeter and Crossover	108°	0.000	800
Ajisai	1986	Laser	50°	0.001	1500

Surface Gravity Data
1° × 1° terrestrial mean gravity anomaly from Ohio State University [*Rapp*, 1986]

or meteorological conditions. The awareness, that the changes in one component of the earth's system are related to changes in the other components, has dictated that the concept of monitoring global dynamic change be a central focus of the Earth Observing System satellites to be launched in 1996 [Butler et al., 1984]. Phenomena such as the El Niño and its impact on global meteorology and Greenhouse effect, with its predicted global warming trend, are areas where the combination of satellite laser ranging, satellite altimetry and precise orbit computation methodology will be applied to yield important observations of the change associated with these phenomena [Salstein and Rosen, 1985].

For example, satellite altimeter observations will be used to monitor short-period and long-period changes in the global oceans [Born et al., 1986]. Prior to the use of satellite altimetry, ocean level was monitored by individual tide gauges, located along the coastlines of continents and islands. Their locations are influenced by tectonic motion and land subsidence, and their relative locations, especially on different continents and islands, are poorly determined. The precise location of these gau-

Table 2. Orbit Accuracy Comparison

Satellite	Epoch	Nominal* Field (rms)	PTGF2 (rms)	PTGF2A (rms)
Starlette (L)	84/01/22	0.67	0.18	0.18
5-day	84/01/27	0.53	0.13	0.15
arcs	85/01/01**	0.60	0.29	0.30
	85/01/06**	0.66	0.21	0.22
Lageos (L)	85/01/01	0.102++	0.097	0.097
30-day	86/01/01**	0.074++	0.069	0.069
arcs				
Nova-1 (D)	84/03/30	0.87	0.50	0.51
12-day				
arcs				
Seasat (L)	78/07/28	1.69	1.04	1.11
6-day (D)		0.85	0.80	0.81
arcs (X)		1.05	0.49	0.50
Ajisai** (L)	86/08/28	0.64	0.39	0.20
5-day	86/09/02	0.51	0.26	0.22
arcs	86/09/07	0.54	0.30	0.20
	86/09/12	0.34	0.20	0.18
Geos-2 (L)	77/04/17	3.31	1.15	1.18
30-day				
arc				

Units for laser (L) and altimeter crossover (X) are in meters; doppler (D) is in cm/sec.
* Nominal fields: GEM-L2 (Lageos); PGS-1331 (Ajisai, Starlette); GEM-10B (GEOS-2, Nova-1); PGS-S4 (Seasat).
**Data set not included in solution for PTGF2 and PTGF2A.
++ Estimate J_2

ges must be known in a terrestrial reference frame if the measurements are to be related to the altimeter measurements of the ocean height. It is essential for monitoring sea level that the origin of the reference system used to specify the coordinates for the tide gauges and the altimeter measurements be located at the earth's mass center. The satellite laser ranging to Lageos has located over 50 stations in a terrestrial reference system whose origin is believed to eoincide with the earth's mass center with an accuracy level greater than 5 cm. Since the seeular change of mean sea level due to global warming is estimated to be between 1 mm and 1 cm/yr, the definition amd maintenance of this terrestrial reference system and the maintenance of the positions of the stations in this reference frame will be an important requirement for future satellite positioning systems.

The monitoring of sea level height using future altimeters will require increased precision in both the ground-based and orbiting SLR and altimetric sytems and in the orbit computation methodology. The orbi-

ting of a laser ranging system and altimeter onboard the Eos platforms will provide the ability to determine the altimeter-derived height measurement, tide gauges and other points relative to the fiducial sites which define the SLR-defined terrestrial reference system. These developments will allow an even richer spectrum of problems to be addressed in the future by the combination of accurate range measurements to and from orbiting satellites in combination with accurate modelling and orbit computation methodology.

Achnowledgments. The author acknowledges numerous discussions with a number of colleagues on many points covered in the previous discussion. Paramount among those were the stimulating and thought-provoking discussion with Professor Giuseppe Colombo. We benefited by his insight into the dynamics of the earth, and planets of the solar system, during his many visits to the University of Texas. His seminars and personal discussions have had a lasting impact on the investigations conducted by both senior faculty and graduate students alike. We will miss those contacts.

APPENDIX A

THE PRIMARY FORCES ON A NEAR-EARTH SATELLITE

The forces which affect the motion of a near-earth satellite can be separated into gravitational forces and surfaces forces.

Gravitational Perturbations

The gravitational perturbations include the gravitational effects of the earth, solid earth tides, lunar, solar and planetary perturbations, ocean tide and general relativistic forces.

Earth Gravity

The perturbing accelerations due to the earth's gravity field, $\bar{G}(t)$, can be expressed as

$$\bar{G}(t) = \nabla U = \nabla (U + \Delta U_s + \Delta U_o) \tag{A.1}$$

where the gradient is defined in a body-fixed coordinate system and where

U is the potential due to the solid-body mass distribution, ΔU_s is the potential change due to solid earth tides, and ΔU_o is the potential change due to ocean tides.

Gravitation Potential for the Solid earth

The perturbing potential function for the earth is generally expressed in terms of a spherical harmonic expansion and is referred to as the geopotential

$$U = \frac{GM}{r}\left\{ -\sum_{l=2}^{\infty} \bar{J}_l \left(\frac{a_e}{r}\right)^l \bar{P}_l(\sin\phi) + \sum_{l=2}^{\infty} \sum_{m=1}^{l} \left(\frac{a_e}{r}\right)^l \bar{P}_{lm}(\sin\phi)\left[\bar{C}_{lm} cos m\lambda + \bar{S}_{lm} \sin m\lambda\right]\right\} \tag{A.2}$$

where the origin of the spherical coordinates is assumed to coincide with the center of mass of the earth (hence, $J_1 = C_{11} = S_{11} = 0$) and $\bar{P}_l$ ($\sin\phi$) represents the degree l Legendre polynomials, $\bar{P}_{lm}$ ($\sin\phi$) represents the associated functions of Legendre of degree l and order m, a_e is the mean equatorial radius of the earth, $\bar{J}_l = -C_{lo}$ represent the normalized zonal harmonic coefficients, and $\bar{C}_{lm}$, $\bar{S}_{lm}$ represent either the normalized tesseral harmonic coefficients, if $l \neq$ m, or the normalized sectoral harmonic coefficients, if $l = m$. The normalized spherical harmonic coefficients are functions of the earth's mass distribution and are obtained from either observation of the motions of artificial satellites or surface measurements of the gravitational acceleration. For any given satellite, the summations over l in Eq. (A.2) are taken to a finite number, n_s, rather than ∞. The dominant perturbation is produced by the earth's oblateness, represented by the geopotential coefficient, J_2.

Solid Earth Tides

Since the earth is not perfectly rigid, it deforms under the solar and lunar gravitational attractions. These deformations, associated with the redistribution of mass, can be defined by using the representation introduced by A. Love in 1909. The deformation can be expressed as a change to the external geopotential by the following expression

$$\Delta U_s(r) = \sum_{l=2}^{\infty} k_l \left(\frac{R_e}{r}\right)^{2l+1} V_l(r) \tag{A.3}$$

where $\Delta U_s(r)$ is the change in potential at position r, k_l are Love numbers of degree l, and V_l is the disturbing tidal potential of degree 1.

The luni-solar tidal potential $V_l(r)$ can be expressed in the following form

$$V_l(r) = GM_i \left(\frac{1}{\varrho_i} - \frac{\bar{r}_i \cdot \bar{r}}{r_i^3}\right) \tag{A.4}$$

where M_i is the mass of the i th disturbing body, the subscript i refers to either the moon or the sun; ϱ_i is the distance between the ith disturbing body and a particle on the solid earth; and r_i is the position vector of the i^{th} disturbing body.

The changes in geopotential caused by the luni-solar tides can be expressed in terms of time dependent geopotential coefficients; that is, $\Delta\bar{C}_{lm}$ and $\Delta\bar{S}_{lm}$ which can be expressed as follows [Sanchez, 1974],

$$\Delta\bar{C}_{lm} = \frac{k_l R_e^{n+1}}{GM} q_{lm} \left[\frac{4(l+2)(l-m)!}{(l+m)!}\right]^{1/2}$$

$$\Delta\bar{S}_{lm} = \frac{k_l R_e^{n+1}}{GM} u_{lm} \left[\frac{4(l+2)(l-m)!}{(l+m)!}\right]^{1/2}$$

and

$$q_{lm} = \sum_{j=2}^{3} \frac{GM}{r_{1j}^{l+1}} \frac{2(l-m)!}{(l+m)!} \bar{P}_{lm}(\cos\theta_j)\cos m\psi_j$$

$$u_{lm} = \sum_{j=2}^{3} \frac{GM}{r_{1j}^{l+1}} \frac{2(l-m)!}{(l+m)!} \bar{P}_{lm}(\cos\theta_j)\sin m\psi_j$$

where the index $j = 1,2,3$ denotes earth, moon and sun, respectively; (θ_j, ψ_j) are colatitude and longitude of the disturbing body; $\Delta\bar{C}_{lm}$ and $\Delta\bar{S}_{lm}$

are the time-varying geopotential coefficients affected by the luni-solar tidal effect. Furthermore, a parameter δ, which represents tidal lag angle associated with the response of the solid earth to the tidal forces, can be used for the time delay caused by the earth's inelasticity.

Ocean Tides

The dynamical contribution of ocean tides due to the gravitational attraction of the sun and moon can be formulated in terms of time-varying geopotential coefficient corrections. The disturbing ocean tide potential, ΔU_o, can be expressed as follows

$$\Delta U_o = 4\pi G \varrho_w R_e \sum_{\mu} \sum_{l=0}^{\infty} \sum_{m=0}^{l} \sum_{+}^{-} \frac{1+k_l'}{2l+1} \left(\frac{R_e}{r}\right)^{n+1} \bar{P}_{lm}(\sin\phi) \cdot \tilde{C}^{\pm}_{\mu lm} \sin(\bar{\eta}_\mu \cdot \bar{\beta}(t) \pm m\lambda \pm \tilde{\epsilon}^{\pm}_{\mu lm}) \tag{A.5}$$

where G is the gravitation constant; ϱ_w is the mean density of the sea water; R_e is the mean equatorial radius; k_l' is the load deformation coefficient for degree l; m is the order of the coefficient; μ is the ocean tide constituent indez; $\bar{\beta}(t) = [\tau\ s\ h\ p\ N'\ p_1]$ are the Doodson arguments which define lunar and solar ephemeris and t is the time; $\bar{n} = [n_1 n_1 \ldots n_6]$ are integer multipliers of Doodson arguments; $\tilde{C}^{\pm}_{\mu lm}$ are amplitudes of ocean tide constituents; and $\epsilon^{\pm}_{\mu lm}$ are the phase angles.

Eq. (A.5) can be expressed conveniently in the form of a correction to the spherical harmonic coefficients, $\bar{C}_{lm}, \bar{S}_{lm}$ given in Eq. (A.2) [Eanes et al., 1983]. If the following definitions are used

$$\begin{pmatrix} C^{\pm} \\ S^{\pm} \end{pmatrix}_{\mu lm} = \tilde{C}^{\pm}_{\mu lm} \begin{pmatrix} \sin \\ \cos \end{pmatrix} (\bar{\eta}_\mu \cdot \bar{\beta}(t) \pm m\lambda) \tag{A.6}$$

and

$$\begin{pmatrix} A \\ B \end{pmatrix}_{\mu lm} = \begin{pmatrix} C^{+} + C^{-} \\ S^{+} - S^{-} \end{pmatrix}_{\mu lm} \cos(\bar{\eta}_\mu \cdot \bar{\beta}) + \begin{pmatrix} S^{+} + S^{-} \\ C^{-} - C^{+} \end{pmatrix}_{\mu lm} \sin(\bar{\eta}_\mu \cdot \bar{\beta}) \tag{A.7}$$

and

$$\text{(A.8)} \qquad F_{lm} = \frac{4\pi R_e^2 \varrho_w}{M} \left(\frac{(l+m)!}{(n-m)!(2l+1)(2-\delta_{0m})} \right)^{1/2} \left(\frac{1+k_l'}{2l+1} \right)$$

where M is the mass of the earth; $C^{\pm}$, $S^{\pm}$ are ocean tide coefficients; and δ_{0m} *is the Kronecker delta function,* $\delta = 1$ for $m = \delta = 0$, then the correction to the lm coefficient can be obtained from Eqs. (A.7) and (A.8) as

$$\text{(A.9)} \qquad \Delta C_{lm} = F_{lm} \sum_{\mu} A_{\mu lm}, \quad \Delta S_{lm} = F_{lm} \sum_{\mu} B_{\mu lm}$$

Resultant Earth Gravitation Potential

As a consequence of Eqs. (A.2), (A.3), (A.5), (A.6), (A.7), (A.8) and (A.9), the resultant gravitation potential for the earth can be expressed as:

$$\text{(A.10)} \qquad u = \frac{GM}{r} \sum_{l=0}^{\infty} \sum_{m=0}^{l} \left(\frac{R_e}{r} \right)^l \bar{P}_{lm}(\sin\phi) \, [(\bar{C}_{lm} + F_{lm} \sum_{\mu} A_{\mu lm}) \cos m\lambda + (\bar{S}_{lm} + F_{lm} \sum_{\mu} B_{\mu lm}) \sin m\lambda].$$

N-Body

The perturbing forces of the sun, moon and other planets, namely, Mercury, Venus, Mars, Jupiter, Saturn, Uranus, Neptune and Pluto, can be approximated with a sufficient accuracy as a point masses. In the nonrotating geocentric coordinate system, the central body and the N-body forces can be expressed as

$$\text{(A.11)} \qquad \bar{P} = \sum_{\mu} CM_i \left[\frac{\bar{r}_i}{r_i^3} - \frac{\bar{\Delta}_i}{\Delta_i^3} \right]$$

where $\bar{P}$ is the force acting on the satellite due to the attraction of the nth body. M is the mass of the earth, $\bar{r}_i$ is the position of the ith mass

with respect to the nonrotating geocentric reference frame, and $\overline{\Delta}_i$ is the position vector between the satellite and the perturbing mass M_i.

The values of $\bar{r}_i$ can be obtained using the planetary ephemerides, for examples, the Jet Propulsion Laboratory Development Ephemeris-200 (JPL DE-200).

General Relativity

The dynamical effects of general relativity on an earth-orbiting satellite are small, but detectable. The primary effect of relativity on the satellite dynamics in that due to the central body mass. The expression for this effect is

$$\bar{R} = \frac{GM}{c^2 r^3}\left\{\left[2(\beta+\gamma)\frac{GM}{r} - \gamma(\dot{\bar{r}}\cdot\dot{\bar{r}})\right]\bar{r} + [2(1+\gamma)(\bar{r}\cdot\dot{\bar{r}})]\,\dot{\bar{r}}\right\} \tag{A.12}$$

where c is the speed of light in vacuum, and M is the mass of the central body, and β and γ are the PPN metric parameters, whose values are usually one.

For circular satellite motion, Eq. (A.12) reduces to

$$\bar{R} = \frac{3(GM)^2\bar{r}}{c^2 r^4}\,. \tag{A.13}$$

From Eq. (A.12), the relativistic perturbation is on the order of $1/c^2$ and is dominantly in the radial direction. In fact, the relativistic perturbation increases slightly the effective GM of the earth. While the central body relativity effect is the dominant effect for near-earth satellites, other effects such as the geodesic precession and the Lense-Thirring effect are important dynamical effects in some application [Ries et al., 1988].

Nongravitational Perturbations

In addition to the gravitational forces due to the earth, moon and planets, other forces which act on the surface of the satellite must be modeled. The primary surface forces are drag and the effects of direct solar radiation and radiation reflected and emitted by the earth.

Atmospheric Drag

The dominant feature of atmospheric resistance for most satellites is a drag force in the direction opposite to the rclative wind. The drag acceleration is usually modeled as

$$\bar{D} = -\frac{1}{2}\varrho(h)\left(\frac{C_D A}{m}\right) V_r^2 \bar{u} \tag{A.14}$$

where ϱ is the atmospheric density, h is the altitude of the satellite above the earth, C_D is the drag coefficient, $\bar{V}_r$ is the velocity of the satellite relative to the atmosphere, A is the cross-sectional area perpendicular to $\bar{V}_r$ m is the satellite mass, V_r is the magnitude of $\bar{V}_r$. and $\bar{u}$ is a unit vector in the $\bar{V}_r$ direction. The drag force is considerably larger than lift forces which, if they exist, would be perpendicular to $\bar{V}_r$. The parameter $(C_D A/m)$ is referred to as the ballistic coefficient, β and A/m is the area-to-mass-ratio. The drag coefficient, C_D is a function of the geometry of the satellite and the Mach number, e.g., the ratio of the vehicle speed to the speed of sound. The parameters, C_D and A, will change if the satellite orientation changes, producing time variations in β. The density $\varrho(h)$ is a complicated function, usually represented by complicated models fit to measurements of the atmospheric properties and its effects on a few satellites within a restrictive flight regime. The density varies with time as a function of the solar flux and the current planetary geomagnetic activity [Jacchia, 1971; Barlier et al., 1977]. Drag has an important effect on the orbit, producing a secular change in the semimajor axis and the eccentricity which causes the orbit to decay, i.e., the orbit tends to spiral inward with the perigee attitude becoming smallcr.

From the data collected on motions of satellites and *in situ* measurements, some basic characteristics of the upper atmosphere density have emerged. These characteristics include:

1. A nearly diurnal variation produced by the solar heating. The subsolar region can be modeled with an atmospheric bulge, the axis of which approximately lags the earth-sun line by about two hours.
2. Solar activity has an important influence through disturbances caused by solar flares and solar plasma events. The frequencies of these phenomena follow the eleven-year solar cycle and the 27-day solar rotation period.
3. The density is influenced by geomagnetic activity and interaction with the charged particles in the upper atmosphere. This phenomena can

produce significant changes in the density at time scales of a few hours to a day.

4. Seasonal changes in the density, including both annual and semi-annual variations, have been observed and are included in current models.

Most of these phenomena cannot be predicted accurately in advance of their occurrence. The consequence of this difficulty is that considerable uncertainty exists with long-term prediction of satellite lifetimes and that precise orbit computations can only be performed after the fact, i.e., after data have been collected on solar and developed geomagnetic activity.

Solar Radiation Pressure

Radiation from the sun produces a small force by transferring momentum from particles streaming out from the sun to a satellite. This force a spherical symmetric satellite is approximately

$$\bar{S} = -P(1+\eta)\frac{A}{M}\nu\,\bar{u} \tag{A.15}$$

where $\bar{F}$ is the direct solar radiation force per unit mass, P is the momentum flux due to the sun, A is the cross-sectional area of the satellite normal to the sun, $\bar{u}$ is the unit vector pointing from the satellite to the sun, η is the reflectivity coefficient with values between 0 and 1, and ν is an eclipse factor such that $\nu = 0$ if the satellite is in shadow of the sun, $\nu = 1$ if the satellite is in sunlight, $0 < \nu < 1$ if the satellite is in partial shadow or penumbra. The passage of the satellite from sunlight to shadow is not abrupt, but the interval of time spent in partial shadow will be very brief for near-earth satellites.

A simple cylindrical shadow model can be used to determine the eclipse factor. Consider the sun forms a cylindrical shadow region behind the earth; the position of the satellite can thus be computed and determined whether it is in the sunlight or in complete shadow.

For composite satellites that are not spherically symmetric and carry rotating solar panels or other antenna arrays, Eq. (A.15) needs to be modified to include area variation effects as in the case of the drag force.

Earth Radiation Pressure

For a close-earth satellite, the solar radiation pressure perturbation due to sunlight reflected from the earth is sometimes not negligible. The radiation pressure of a spherically symmetric satellite due to the earth albedo can be expressed in the form of Eq. (A.15) as follows

$$\text{(A.16)} \qquad \bar{E} = -P'(1+\eta)\frac{A}{M}\nu\bar{r}$$

where $\bar{F}$ is the earth albedo radiation pressure per unit mass, ν is the eclipse factor, P' is the solar momentum flux due to the earth albedo, and $\bar{r}$ is the radial position vector normal to the effective reflecting area of the earth to the satellite. Eq. (A.16) is highly simplified, since it treats the effective reflecting surface as a reflecting spherical disc and assumes reflective property of the surface to be the same. P' is a function of the earth albedo, γ, and the angle of incidence of the reflected sunlight. γ can be approximated within five-percent precision by a five-parameter model given by

$$\text{(A.17)} \qquad \gamma(\phi) = a_o + a_2\sin^2(\phi + c_2) + a_4\sin^4(\phi + c_4)$$

where ϕ is the geodetic latitude; a_0, a_2, a_4 are arbitrary constants; and c_2 and c_4 are phase angles.

Eq. (A.16) can be modified to take into account the two factors for more accurate representation of the earth albedo model. First, the radiation model can be better formulated to account separately for the flux due to optical and infrared radiation. Second, the earth's surface can be subdivided into segments in a finite element approach, and the momentum flux can be properly integrated over each surface area to obtain a better representation of the earth albedo [Knocke, 1989].

Again, the assumption made in arriving at Eq. (A.16) is that the satellite is spherically symmetric. Note that the surface area, A, and the reflectivity coefficient, η, can be different from that of the direct solar radiation pressure in Eq. (A.15).

REFERENCES

APEL, J.R., Satellite sensing of ocean surface dynamics, *Ann. Rev. Earth Planet Science, 8*, 303-342, 1980.

BARLIER, F., BERGER, C., FALIN, J.L., KOCKARTS, G. and THUILLIER, G., A thermospheric model based on satellite drag data, *Aeronomica Acta, 185*, 1977.

BORN, G.H., TAPLEY, B.D., RIES, J.C. and STEWART, R.H., Accurate measurement of mean sea level changes by altimetric satellites, *J. Geophys. Res., 91* (C10), 11,775-11,782, Octobert 15,1986.

BUTLER, D.M. et al., Earth observing system: Science and mission requirements working group report, Vol. 1, *NASA TM-86129*, 1984.

CHAO, B.F., Interannual length-of-day variation with relation to the southern oscillation/El Niño, *Geophys. Res. Lett., 11*, 541-544, 1984.

CLARK, T.A., GORDON, D., HIMWICH, W.E., MA, C., MALLAMA, A. and RYAN, J.M., Determination of relative site motions in the western United States using Mark III VLBI, *J. Geophys. Res., 92,* 12, 741-12,750, 1987.

COATES, R.J., FREY, H., BOSWORTH, J. and MEAD, G., Space-age geodesy: The NASA crustal dynamics project, *IEEE Trans. on Geoscience and Remote Sensing, GE-23* (4), 358-368, July 1985.

COUNSELMAN, C.C., and STEINBRICHER, D.H., Miniature interferometric terminal for earth surveying: Ambiguity and multipath with Global Positioning System, *IEEE Trans. Geosci. and Remote Sensing, GE-19,* 224-252, 1981.

EANES, R.J., SCHUTZ, B.E. and TAPLEY, B.D., Earth and ocean tide effects on Lageos and Starlette, *Proc. Ninth International Symposium on Earth Tides*, J.T. Kuo, Ed., E. Schweirzerbart'she Verlagsbuchhandlung, 1983.

FISHER, R.A., On an absolute criterion for fitting frequency curves, *Messenger of Math, 41,* p. 155, 1912.

GAUSS, K.F., *Theoria Motus*, 1809; also, *Theory of the Motion of the Heavenly Bodies About the Sun in Conic Sections*, Dover, New York, 1963.

JACCHIA, L., Revised static models of the thermosphere and exosphere with empirical temperature profiles, *Special Report 332,* Smithsonian Astrophysical Observatory, Cambridge, Mass., 1971.

JOHNSON, C.W., LUNDQUIST, C.A. and ZURASKY, J.L., The Lageos Satellite, International Astronautical Federation XXVIIth Congress, Anaheim, Calif., October 1976.

KNOCKE, P.C., Earth radiation effects on satellites, *CSR-89-1*, Center for Space Research, The University of Texas at Austin, January 1989.

KOLMOGOROV, A.N., Interpolation und Extrapolation von stationaren zufalligen Folgen, *Bull. Acad. Sci. USSR*, Ser. Math. 5, 3-14, 1941.

LAWSON, C.L. and HANSON, R.J., *Solving Least Squares Problems*, Prentice Hall, 1963.

LIEBELT, P.B., *An Introduction to Optimal Estimation*, Addison-Wesley, 1967.

MARSH, J.G. et al., A new gravitational model for the earth from satellite tracking data: GEM-T1, *J. Geophys. Res., 93*(B6), 6169-6215, June 1988.

MAYBECK, P.S., *Stochastic Models, Estimation and Control, 1*, Academic Press, 1979.

MELBOURNE, W., ANDERLE, R., FEISSEL, M., KING, R., MCCARTHY, D.M., SMITH, D., TAPLEY, B.D. and VICENTE, R., *Project MERIT Standards*, U.S. Naval Observatory Circular No. 167, Washington, D.C., December 27, 1983.

PLOTKIN, H.H., JOHNSON, T.S. and MINOTT, P.O., Progress in laser ranging to satellites: Achievements and plans, *The Use of Artificial Satellites for Geodesy and Geodynamics*, G. Veis, Ed., National Technical University of Athens, 1973.

RIES, J.C., HUANG, C. and WATKSIN, M.M., Effect of general relativity on a near-earth satellite in the geocentric and barycentric reference frames, *Phys. Rev. Lett., 61*(8), 903-906, August 22, 1988.

ROBERTSON, D.S., CARTER, W.E., TAPLEY, B.D., SCHUTZ, B.E. and EANES, R.J., Polar motion measurements: Subdecimeter accuracy verified by intercomparison, *Science, 229*, 1259-1261, September 1985.

ROSBOROUGH, G.W., Satellite orbit perturbations due to the geopotential, *CSR-86-1*, Center for Space Research, The University of Texas at Austin, 1986.

ROSBOROUGH, G.W. and TAPLEY, B.D., Radial, transverse and normal satellite position perturbations due to the geopotential, *Celest. Mech., 40*(3-4), 1987.

RUBINCAM, D.P., Postglacial rebound observed by Lageos and the effective viscosity of the lower mantle, *J. Geophys. Res., 89*, 1077-1087, 1984.

SALSTEIN, D.A. and ROSEN, R.D., Earth rotation data as a proxy index of global wind fluctuations, Third Conference on Climate Variations and Symposium on Contemporary Climate, 1850-2100, 1985.

SANCHEZ, B.V., *Rotational Dynamics of Mathematical Models of the Nonrigid Earth*, Ph.D. dissertation, The University of Texas at Austin, August 1974.

SMITH, D.E., CHRISTODOULIDIS, D.C., KOLENKIEWICZ, R., DUNN, P.J., KLOSKO, S.M., TORRENCE, M.H., FRICKE, S. and BLACKWELL, S., A global geodetic reference frame from Lageos ranging (SL5.1AP), *J. Geophys. Res., 90* (B11), 9221-9234, September 1985.

SORENSON, H.W., Least-squares estimation: From Gauss to Kalman, *IEEE Spectrum,* 63-68, July, 1970.

STEWART, R.H., FU, L.L. and LEFEBVRE, M., Science opportunities from the Topex/Poseidon Mission, *JPL Publ. 86-18*, NASA Jet Propulsion Laboratory, Pasadena, Calif., July 1986.

TAPLEY, B.D., Statistical Orbit Determination Theory, *Recent Advances in Dynamical Astronomy,* D. Reidel Publishing Co., 396-425, 1973.

TAPLEY, B.D., and BORN, G.H., The Seasat precision orbit determination experiment, *J. Astron. Sci., XXVIII*(4), 315-326, 1980.

TAPLEY, B.D., BORN, G.H. and PARKE, M.E., The Seasat altimeter data and its accuracy assessment, *J. Geophys. Res., 87*(C5), 3179-3188, April 30, 1982.

TAPLEY, B.D., Polar motion and earth rotation, *Rev. Geophys., 21*(3), 569-573, 1983.

TAPLEY, B.D., SCHUTZ, B.E. and EANES, R.J., Satellite laser ranging and its applications, *Celestial Mechanics, 37*, 247-261, 1985.

TAPLEY, B.D., SCHUTZ, B.E. and EANES, R.J., Station coordinates, baselines and earth rotation from Lageos laser ranging: 1976-1984, *J. Geophys. Res., 90*(B.11), 9235-9248, September 30,1985.

TAPLEY, B.D. and ROSBOROUGH, G.W., Geographically correlated orbit error and its effect on satellite altimetry missions, *J. Geophys. Res., 90*(C6), 11817-11831, November 1985.

TAPLEY, B.D., SHUM, C.K., YUAN, D.N., RIES, J.C. and SCHUTZ, B.E., An improved model for the earth's gravity field, *Proc. Chapman Conference on Progress in the Determination of the Earth's Gravity Field*, September 1988.

Topex Science Working Group, Satellite Altimetric Measurements of the Ocean, *No. 400-111*, Jet Propulsion Laboratory, Pasadena, Calif., March 1, 1981.

WIENER, N., *The Extrapolation, Interpolation and Smoothing of Stationary Time Series*, John Wiley & Sons, Inc., New York, 1949.

WILKINS, G.A. (Ed.), *Project MERIT: A Review of the Techniques to Be Used During Project MERIT to Monitor the Rotation of the Earth*, p. 77, Royal Greenwich Observatory, Herstmonceux, 1980.

YODER, C.F., WILLIAMSON, J.C., DICKEY, J.O., TAPLEY, B.D., EANES, R.J. and SCHUTZ, B.E., Secular variations of the earth's gravitation harmonic J_2 coefficient from Lageos and nontidal acceleration of earth rotation, *Nature, 303*(5920), 757-762, June 30, 1983.

Misure di precisione dell'orbita di un satellite artificiale

Sigfrido LESCHIUTTA *

Sommario. *La posizione e l'orientamento nello spazio di un satellite artificiale o di una sonda spaziale devono essere noti e regolati da Terra, con imprecisioni varie, che dipendono ovviamente dalla missione che il dispositivo deve svolgere. Nei primi trenta anni di esplorazioni spaziali, la imprecisione con la quale può essere ricavata l'orbita è scesa di oltre cinque ordini di grandezza, passando da alcuni kilometri a pochi centimetri. Questo progresso è dovuto al fatto che le classiche misure di angoli, proprie della Astronomia, si è passati a misure di distanza o di variazione di distanza, mediante apparati radioelettrici, basati su orologi atomici.*

Scopo di questa nota è il presentare la situazione e le prospettive di questo tipo di ricerche unicamente per la posizione, rinviando ad altra sede per quanto riguarda l'orientamento.

1. Introduzione

La determinazione dell'orbita di un satellite artificiale o di una sonda spaziale viene effettuata misurando, in funzione del tempo, posizioni successive dell'oggetto, espresse rispetto un opportuno sistema di riferimento. Queste posizioni sono ottenute, come si vedrà, mediante tutta una serie di metodi radioelettrici, che vengono confrontati cronologicamente nella fig. 1, sulla quale si tornerà più oltre; in ordinate si ha l'imprecisione con la quale è possibile determinare il semiasse maggiore dell'orbita, o la distanza satellite - stazione terrestre.

Dalla figura risulta che attualmente queste misure sono possibili con errori dell'ordine di qualche centimetro, mentre attività sono in corso per raggiungere, tra qualche anno, l'imprecisione di 5 mm; tre-quattro cm sull'orbita di un satellite geostazionario, equivalgono ad imprecisioni dell'ordine di 10^{-9}.

È opportuno esaminare, sia pure brevemente, i requisiti di precisione per posizione ed orientamento posti dalle varie classi di satelliti; questo

* Politecnico di Torino - Dipartimento di Elettronica.

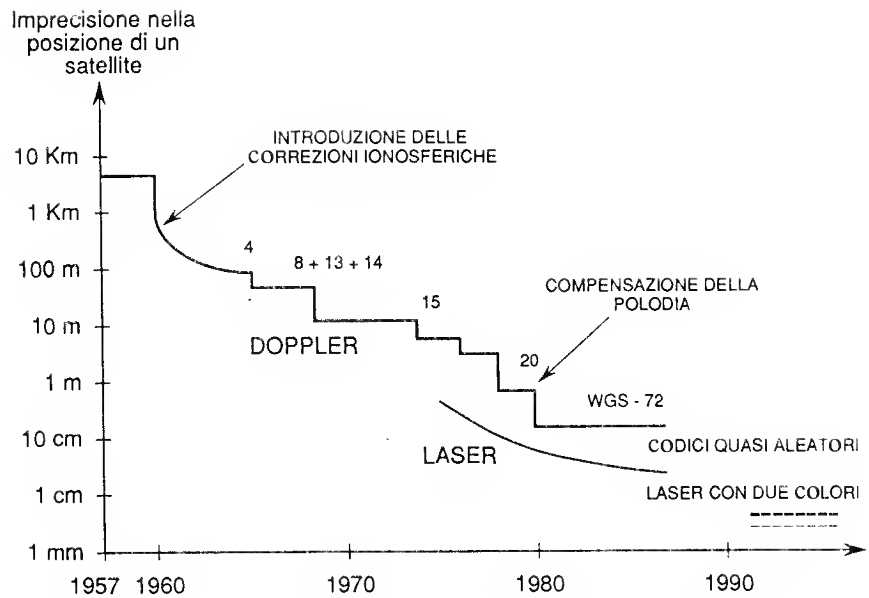


Fig. 1 - Andamento della imprecisione nella misura della posizione di un satellite in funzione del tempo e di cinque metodi spermentali. Per il metodo Doppler sono indicati i miglioramenti conseguiti con la introduzione delle correzioni dovute agli effetti ionosferici ed alla polodia; i numeri si riferiscono all'armonico più elevato del geopotenziale usato nei calcoli.

panorama viene presentato, prevalentemente sotto forma di tabelle, nella seconda sezione della nota.

Questa nota è dedicata ai metodi sperimentali necessari per ricavare i dati che verranno utilizzati per il caleolo di un'orbita o traiettoria, ma, data la sede nella quale il lavoro viene presentato, con la presenza di studenti di Università e Politecnico, può essere opportuno un richiamo introduttivo svolto nella terza sezione, sulla parte teorica e sulle complessità di conoscenze e di calcolo necessarie per ricavare l'orbita di un satellite.

Nella quarta sezione, dopo alcune considerazioni sul ruolo essenziale delle misure di tempo e di frequenza in queste ricerche, vengono esposti partitamente i tre metodi fondamentali usati per ricavare i dati necessari per calcolare le posizioni e quindi l'orbita di un satellite artificiale, indicando per ognuno di questi le varianti e gli eventuali sistemi esistenti.

I possibili sviluppi di questo tipo di ricerche, con particolare interesse al ruolo dell'Agenzia Spaziale Europea, ESA, e di quella italiana, ASI, verranno esposti nella quinta sezione, mentre nella sesta ed ultima ven-

gono elencati alcuni dei risultati, molti dei quali inaspettati, che la misura della posizione di una sonda o di un satellite hanno eonsentito ad altre discipline, dalla Tettonica a placche alla Astronomia di posizione.

Resta da spiegare perché in una giornata di studi dedicata alla memoria di Giuseppe Colombo, si parli dei metodi sperimentali usati per la determinazione di orbite.

Giuseppe Colombo era un meccanico celeste ed un ingegnere, e quindi è naturale che, da quando l'Astronomia divenne una scienza sperimentale alla portata dell'uomo, nascesse in Lui un interesse diretto per lo studio, in tutti i suoi aspetti, di quel curioso sistema planetario costruito attorno alla Terra con il lancio di alcune migliaia di satelliti; da ingegnere, la sua «curiositas» non si fermava a metodi e risultati, ma si spingeva anche alle tecniche usate.

Colombo, in particolare, ha stimolato la realizzazione in Italia di un satellite dedicato allo studio di orbite, il LAGEOS 2; l'oggetto, già costruito, attendc il laneio.

Colombo ha promosso la costruzione in Italia di una stazione di telemetria laser per lo studio di orbite, entrata in servizio nel 1986.

Colombo ha direttamente ed indirettamente stimolato molti giovani e meno giovani ricercatori italiani ad interessarsi al problema del calcolo delle orbite e all'uso di questi dati per le più disparate ricerche, dalla relatività alla tettonica a placche.

Bcpi, infine, sin dai lontani progetti dei due satelliti che si inseguono nello spazio sia dalla idea della «freccia verso il Sole», con a bordo un maser all'idrogeno, ha stimolato e coinvolto nelle tecniche spaziali, con le suc tipiche telefonate che, partendo dai continenti più disparati arrivavano a Torino sempre all'ora canonica, chi ha preparato queste note.

2. Requisiti di precisione per l'orbita dei satelliti

Satelliti o sonde vengono lanciati e posti in orbite o traiettorie che sono diverse in funzione dello scopo della missione.

Gli scopi più comuni sono:

telecomunicazioni	astronomia	astronomia di posizione
sorveglianza	meteorologia	risorse terrestri
geofisica	radionavigazione	esplorazioni interplanetarie
radiodiffusione	geodesia	oceanografia
aeronomia	soccorso	soccorso in mare

In taluni casi (ad es. telecomunicazioni) il satellite, geostazionario, deve

essere mantenuto entro alcune decine di chilometri dalla posizione assegnata, ma è necessario conoscerne la posizione con errori inferiori al kilometro per poter effettuare, con opportuno anticipo, le operazioni di modifica dell'orbita necessarie per restare in «stazione».

La posizione deve essere mantenuta per tutta una serie di esigenze, non ultime quelle di carattere politico.

In tabella vengono riportati i valori attualmente richiesti e quelli proposti per il futuro.

Seeondo le convenzioni solitamente usate, le coordinate x ed y sono disposte rispettivamente lungo ed in direzione ortogonale all'orbita, formando una terna trirettangola con la coordinata z, che è radiale e che quindi rappresenta l'altezzza del satellite.

I valori indicati tra parentesi corrispondono, ove del caso, alle future normative internazionali.

Tabella I

Satelliti	**Posizione**		**Assetto**
Geostazionari:			
Comunicazioni e meteorologici	+/−0,1° (+/−0,01°)		0,1°-1°
Radiodiffusione	+/−0,1° (0,05° Lat.) (0,02° Lon.)		0,1° rollio 0,05° imbardata 0,1° beccheggio 0,05°
Risorse terrestri	x - y	1 - 10 km	30'' - 1°
	z	0,1 - 20 m	
Oceanografia	x - y	alcuni km	0,1° - 1°
	z	1 - 5 cm	
Gravimetria - geodinamica	x - y - z	1 cm	
Navigazione	x - y - z	0,1 - 30 m	

Solitamente l'orbita deve essere nota per i due motivi seguenti:

— il satellite deve mantenere una certa orbita (ed un certo orientamento) necessaria per lo svolgimento della missione.

— la grandezza di interesse è l'orbita stessa del satellite, in quanto è dallo studio di questa orbita che si ricavano le informazioni cercate; si

richiede pertanto la massima precisione consentita dalla tecnica e dalle modalità sperimentali. I satelliti usati in Geodesia, Astronomia di posizione e per talune ricerche di Geofisica rientrano in queste due ultime categorie.

Nel primo caso i requisiti di precisione sono i più vari; vengono forniti alcuni esempi:

Satellite per oceanografia: scopi siano il rilievo della topografia degli oceani, della altezza delle onde, della intensità e direzione dei venti sulla superficie del mare. Il satellite, munito di particolari sensori radar, segue un'orbita circolare, la cui componente radiale, rispetto alla Terra, deve essere nota con errori inferiori alle imprecisioni consentite per le misure, e solitamente inferiore al metro. Le altre due componenti (lungo ed ortogonale all'orbita) — che individuano la zona dell'oceano studiata, non sono critiche ed è ammesso un errore di 0.1 - 1 km.

L'imprecisione «radiale» deve essere mantenuta anche in mezzo ad un oceano, cioè in punti dell'orbita fuori dalla portata delle stazioni di misura che sono per definizione a Terra.

Satellite per navigazione aerea: ogni errore di posizione si ripercuote direttamente sulla determinazione della posizione, pertanto è necessario che la posizione del satellite sia nota con errori dell'ordine del metro, ed anche minore ove si volesse sfruttare il satellite come «aiuto» per l'atterraggio.

3. Calcolo dell'orbita

I sei parametri kepleriani che individuano il piano orbitale, l'orbita e la posizione del satellite lungo l'orbita sono ricavati mediante misure di distanza, posizione o velocità relativa e tempo, ottenuti mediante i metodi sperimentali indicati nella prossima sezione 4 e seguendo due approcci.

Il primo, ormai classico e ben consolidato, si basa su considerazioni di dinamica, il secondo, non ancora ben sviluppato, in via di introduzione e consentito da taluni nuovi metodi sperimentali, è fondato su semplici considerazioni geometriche.

Il primo metodo, chiamato *dinamico*, considera inizialmente le varie forze che agiscono su un satellite, la cui risultante provoca una accelerazione. Partendo da un gruppo di valori iniziali di posizione assunti per un tempo noto, usando un modello del geopotenziale, si integra due volte la accelerazione sino a coprire l'intervallo di tempo per il quale si dispone di dati sperimentali, ad es. di posizione.

Dal confronto tra le previsioni del modello matematico e le misure, con classiche tecniche ai minimi quadrati, si determinano o le *correzioni* ai valori assunti inizialmente, o altre *incognite*.

Le cose sono ben più complicate perché le forze che agiscono sul satellite sono numerose ed intervengono con meccanismi differenti.

Un elenco non esaustivo deve considerare forze:

— proporzionali alla massa come le forze gravitazionali di Terra, Sole, pianeti; il geopotenziale stesso deve essere sviluppato in una serie di armonici, che arrivano al trentesimo ordine.
— proporzionali alla superficie ed alla forma del satellite, come l'albedo e le varie pressioni di radiazione, in primo luogo quella solare, e le accelerazioni risultanti da reirradiazione dovuta a gradienti termici entro il corpo del satellite, ed, al limite, il rinculo per emissione di ogni forma di energia.
— dipendenti da forma e tipo di superficie, come le interazioni con campi elettrici e magnetici esistenti nello spazio.

Con riferimento alla fig. 1, i numeri riportati lungo la curva che si riferisce al metodo Doppler, indicano il numero dei termini armonici del geopotenziale considerati dal modello usato per il calcolo dell'orbita.

I calcoli sono ulteriormente complicati dal fatto che dovrebbero essere effettuati in un sistema inerziale non ruotante, mentre tutte le misure sono effettuate dalla Terra ed alcune forze agenti sul satellite dipendono anche esse dai vari moti della Terra.

Da qui la *necessità* di conoscere questi moti — polodia, variazioni della durata del giorno ed altri — con adeguata precisione e tempestività, ma anche la *possibilità*, invertendo le equazioni, di ricavare questi moti, assunta un'orbita ideale o corretta per gli altri termini già noti.

Lo sfruttamento di questa possibilità ha consentito rilevanti progressi alla Astronomia di posizione negli scorsi due decenni.

Le misure stesse, oltre che dagli inevitabili errori strumentali, sono perturbate da effetti ionosferici e troposferici, di non agevole compensazione.

È pertanto evidente che un programma di previsione di orbite è estremamente complesso e che la sua gestione è riservata a specialisti del settore; ovviamente il calcolo di tutte le forze e delle complicazioni più sopra indicate si impone quando si vogliano ottenere le massime precisioni, vuoi per l'orbita, vuoi per i parametri che turbano l'orbita.

Altro vantaggio del metodo *dinamico* è la possibilità di usare qualsiasi tipo di misura relativo all'orbita (sono stati usati anche gli angoli di puntamento delle antenne), di poter utilmente combinare dati di tipo diverso, di precisione diversa e soprattutto effettuati in momenti diversi.

Il secondo metodo, detto *geometrico*, valido unicamente per determinare la posizione del satellite e la sua orbita, richiede invece la determinazione simultanea di tre o più distanze da punti noti a Terra. La posizione è dunque ottenuta per tri- o multilaterazione.

Questo approccio, che è in fase sperimentale, è consentito dai nuovi metodi di misura, di cui al punto 4.2.2, basati dall'uso contemporaneo di più codici quasialeatori o di impulsi laser.

Le posizioni del satellite, necessarie per tutti questi calcoli, richiedono a loro volta la *individuazione di un «luogo»* geometrico che definisca la posizione del satellite stesso e la *scelta di metodi radioelettrici* per individuare senza ambiguità e con la necessaria precisione, i parametri che individuano quei luoghi geometrici.

I luoghi usati sono solitamente:

— *sfere*, la posizione risulta dalla «intersezione» di tre o più sfere, delle quali di misura raggio e posizione del centro,

— *iperboloidi*, analoga intersezione; i luoghi vengono individuati conoscendo le posizioni di quattro punti (possono essere quattro posizioni successive dello stesso satellite) e misurando quattro differenze di distanza che danno luogo a tre iperboloidi,

— *coni*, con vertice nella posizione del satellite, per asse la tangente all'orbita e semiapertura calcolata dal rapporto tra velocità relativa e velocità assoluta del satellite,

— *piani*,

— *varie intersezioni dei quattro luoghi predetti.*

4. Metodi sperimentali

Tutti i metodi sperimentali usati per ricavare la traiettoria di un satellite o di una sonda sono basati sulle proprietà di precisione e di stabilità degli orologi atomici.

Le determinazioni mediante mezzi ottici sono attualmente limitate alla localizzazione delle centinaia di oggetti di rilevanti dimensioni — ultimi stadi, schermi protettivi, satelliti non più attivi — che incombrano lo spazio e non sono provvisti di sistemi radioelettrici di localizzazione.

La tabella II riporta le varie grandezze misurate ed il tipo di misurazione.

Tabella II

Grandezze		Tipo di misura
di interesse	misurate	
distanza	misura di fase	modulazione di portanti interferometria
	tempo di volo	impulso in microonda codice quasi aleatorio impulso luminoso
differenze di distanza o apertura di coni	variazione di frequenza	battimenti tra portanti stabili

4.1.1. Misure di fase - «side tone ranging»

Il segnale usato è costituito da una modulazione impressa su una portante in microonda; inviato al satellite, viene da questo rinviato, previa traslazione di frequenza, alla stazione terrestre, ove si confronta la fase della modulazione di ritorno con la fase della modulazione di partenza.

Il ritardo misurato è la somma dei tempi di propagazione nei due versi (compresi effetti troposferici, ionosferici e relativistici) e di tutti i ritardi strumentali a Terra ed entro il satellite. Questi ritardi ed effetti possono essere valutati o misurati. Depurando il risultato della misura del ritardo, la distanza voluta è ottenuta semplicemente moltiplicando per c la metà del ritardo.

Questo metodo, estremamente semplice, conosciuto con il nome di «*side tone ranging*», è usato per la determinazione della posizione delle centinaia di satelliti geostazionari che assicurano i servizi di telecomunicazione, radiodiffusione dallo spazio e meteorologici, e verrà descritto con un certo dettaglio.

Le operazioni vengono avviate, adottate una frequenza di modulazione sufficientemente «bassa» per evitare le ambiguità risultanti alla presenza nel tragitto «Terra-Satellite-Terra di un numero, a priori sconosciuto, di periodi corrispondenti alla frequenza di modulazione.

Il criterio seguito è quello di cominciare con una frequenza di modulazione tale che la corrispondente lunghezza d'onda sia maggiore, per almeno un fattore due, all'ambiguità con la quale sia già nota, per altra via, la distanza del satellite.

Il valore *due* risulta dalla considerazione che il segnale percorre due volte la stessa distanza.

Si supponga, fig. 2, che si sappia che la distanza sia compresa tra 6000 e 6500 km con un ulteriore errore di ± 1000 km. Si addotti una frequenza di modulazione di 50 Hz., cui corrisponde un periodo di 20 ms ed una lunghezza d'onda di 6000 km e si osservi, al ritorno, una differenza di fase $\varphi = 45$ gradi.

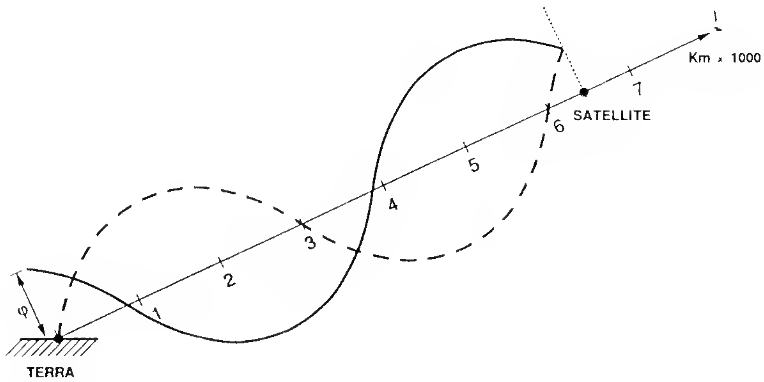


Fig. 2 - Calcolo della distanza tramite la misura della fase di una modulazione; per i valori numerici si veda il testo.

La distanza vale $d = c\,(\varphi + kT)\,/\,2$ con k intero, corrispondente al numero di lunghezze d'onda presenti lungo il collegamento. È immediato constatare che solo il valore $k = 2$ da un valore compreso tra 5000 e 7500 km.

Una misurazione di fase viene eseguita con errori residui attorno un grado; ammettendo che l'errore sia di 2 gradi, risulta una incertezza di ± 15 km sulla distanza stazione-satellite.

Sostituendo, si ricava che la distanza è di 6375 ± 15 km.

Noto un primo valore della distanza, la frequenza di modulazione viene progressivamente elevata in modo da ridurre l'errore di risoluzione; le ambiguità vengono risolte, in materia automatica, con una opportuna scelta delle frequenze di modulazione.

Non esistono limiti teorici alla risoluzione del metodo; infatti, al limite si può passare ad effettuare misure di fase direttamente sulla portante; con questa soluzione, la risoluzione è stata portata ad alcuni millimetri (si presti attenzione al fatto che si parla di risoluzione della misura, non di imprecisione della distanza). Nascono infatti limiti do-

vuti ad altri fattori, quali la propagazione del segnale entro la ionosfera; comunque errori dell'ordine di decine-centinaia di metri per un satellite geostazionario, sono accettabili.

Nell'uso scientifico del metodo, con misure di fase effettuate direttamente sulle portanti e non sulle modulanti, è possibile raggiungere il valore del centimetro, come risoluzione di misura e di alcuni centimetri come *imprecisione*, ricorrendo a particolari tecniche.

4.1.2. Interferometria

Del tutto analoghe alla radiointerferometria a grande base (VLBI - Very Large Base Interferometry), usata in radioastronomia, sono le determinazioni di posizione per sonde spaziali interplanetarie, o per scopi di geodesia di precisione, che ricorrono a tecniche interferometriche.

Due antenne, fig. 3, determinano la differenza di fase (o, il che equivale, di tempo di arrivo) di un segnale proveniente dal satellite.

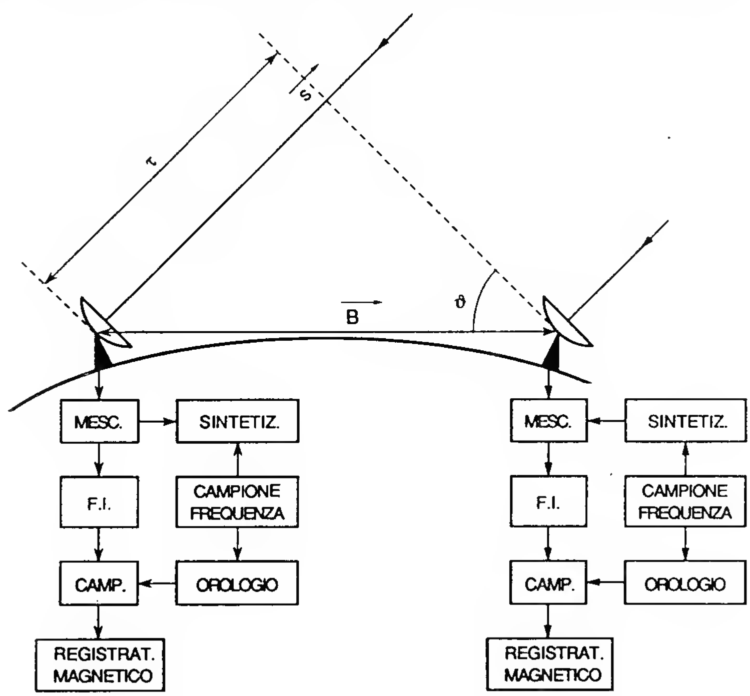


Fig. 3 - Misura della differenza dei tempi di arrivo di un'onda piana in due punti diversi — interferometria con grande base. Il segnale considerato può essere il rumore di una radiostella oppure il segnale proveniente da un satellite.

Per localizzazione di sonde interplanetarie le stazioni a Terra sono poste lontane, su continenti diversi e pertanto non è concepibile di usare, come sarebbe conveniente, lo stesso oscillatore per le misure di fase. Si è pertanto costretti a ricorrere ad orologi astronomici, quali il maser con Idrogeno, entro il quale la transizione a 21 cm dell'idrogeno genera un segnale campione.

Questo tipo di orologio offre una stabilità dell'ordine di $10^{-14}-10^{-15}$ per la durata della misurazione.

Con questo metodo sono state misurate da Terra le posizioni delle sonde interplanerarie che hanno visitato i satelliti di Giove; si hanno errori di poche decine di kilometri ad una distanza di alcune U.A. dalla Terra.

Lo stesso metodo è inoltre usato per determinare le posizioni relative di due punti a terra, effettuando misure di fase sulle portanti emesse dai satelliti di un sistema di navigazione. Con correzioni ionosferiche ed altre troposferiche, basate su misure di pressione, temperatura, grado igrometrico dell'aria presso le due stazioni terrestri, è possibile determinare le posizioni relative con errori residui di qualche centimetro, per distanze da alcune decine a poche centinaia di kilometri.

4.2. Misure di tempo di volo

Il metodo più semplice ed intuitivo per determinare con metodi radioelettrici una distanza è quello di inviare un impulso elettromagnetico al satellite munito di un ricetrasmettitore, mediante il quale il segnale ricevuto viene rinviato a Terra. La stazione determina il tempo di volo totale, dovuto alla propagazione elettromagnetica ed ai ritardi strumentali a Terra e sul satellite. Anche in questo caso si avranno ritardi ionosferici, troposferici ed effetti relativistivi. I ritardi possono essere valutati o misurati e quindi, dopo ovvii calcoli, si ottiene la distanza.

I problemi sorgono alla scelta del segnale da inviare, delle apparecchiature, della banda di frequenza necessaria e della frequenza portante del collegamento.

Inoltre, poiché la funzione di «posizione» non è lo scopo principale del satellite e quindi non si può sempre pensare di imbarcare apposite apparecchiature, con conseguenti aumenti di massa e di consumo di energia, si dovrà cercare di sfruttare per la funzione di «posizione» qualche altro apparato comunque esistente a bordo.

4.2.1. Impulso radioelettrico

Nel caso di un impulso, è evidente che il fronte di salita deve essere «ripido» per ottenere adeguata precisione. Un fronte d'onda con pendenze dell'ordine di 1 ms, darebbe luogo ad errori di 150 km sulla distanza.

Quindi la risoluzione della misura dipende dal tempo di salita del segnale; esiste la relazione $B \cdot t = 0{,}35$, dove B è la banda, in Hz, del collegamento, e quindi di tutte le apparecchiature interessate, t è il tempo di salita, definito con l'intervallo di tempo necessario perché l'impulso passi dal 10% al 90% della propria ampiezza. Ove si volesse un tempo di salita di 10 ns, la banda necessaria sarebbe di 35 Mhz. Detta banda sarebbe inoltre occupata in esclusività dal segnale usato nella determinazione.

La cadenza tra gli impulsi deve inoltre rispettare i criteri di ambiguità e quindi di distanza tra i singoli impulsi, già indicati al punto 4.1.1; per un satellite geostazionario, con tempo di volo totale attorno a 270 ms di potrebbe trasmettere non più di due impulsi al secondo.

Il radiotrasmettitore usati sia a Terra, sia sul satellite, risulterebbero male utilizzati, in quanto sarebbero necessari apparati di rilevante potenza, larga banda ed effettivamente attivi per una estremamente ridotta parte del tempo.

Non essendo praticabile questa strada, si usano solitamente i canali e possibilmente segnali già esistenti per altri motivi e che abbiano le caratteristiche opportune; particolarmente adatti risultano i satelliti per collegamenti televisivi o per la trasmissione diretta di immagini dallo spazio, soprattutto perché i canali disponibili hanno larghezze di banda adeguate.

Con questa soluzione è possibile ricavare la distanza di un satellite geostazionario con risoluzione di 3-5 m e, previa calibrazione dei ritardi strumentali e valutazione o misura di quelli ionosferici, con imprecisione di 15-30 m.

Ove sia possibile una scelta, la frequenza della portante sia la più alta possibile, infatti, a parità di condizioni della ionosfera, i ritardi scendono con il quadrato della frequenza della portante.

4.2.2. Codici quasi-aleatori

Una interessante alternativa all'uso di impulsi consiste nella trasmissione continua, con livello di potenza estremamente ridotto, di un codice binario, che abbia le caratteristiche generali di un rumore bianco, ma

che si ripeta ad opportuni intervalli. Un siffatto segnale può essere *sovrapposto* al segnale principale (immagine video, canali telefonici, ecc.) producendo una degradazione accettabile.

L'intervallo di ripetizione del codice, che viene chiamato quasi aleatorio, in quanto ha le apparenze di una successione casuale di «uno» e «zero», ma non lo è completamente perché ogni tanto si ripete, viene prescelto con gli stessi criteri adottati per eliminare la ambiguità nella misura di distanza.

Con le misure di fase, punto 4.1, si sceglieva un opportuno valore per la frequenza di modulazione, con la trasmissione di impulsi, punto 4.2.1, si sceglieva l'intervallo di tempo tra i singoli impulsi, qui si sceglie la durata del codice.

La misura della distanza viene effettuata, fig. 4, inviando al satellite il codice, sotto forma di una modulazione di una portante, che viene rinviata a Terra, su altra portante.

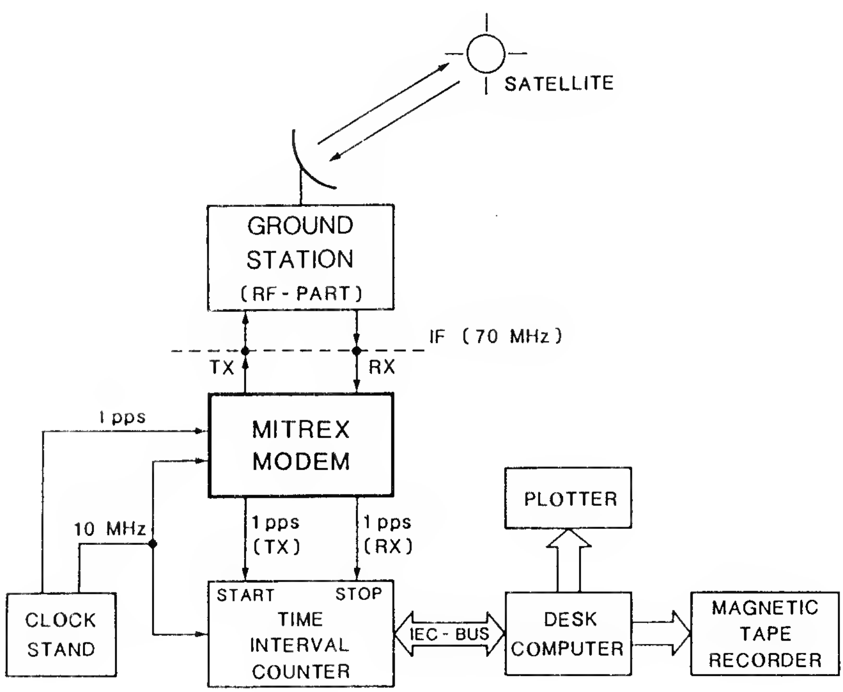


Fig. 4 - Metodo mediante codici quasi aleatori, le operazioni di generazione del codice, ritardo della versione locale ed infine correlazione con il codice che è transitato attraverso il satellite vengono effettuate nell'apparato chiamato MITREX (Microwave Timing and Ranging Experiment).

Il codice ricevuto viene avviato ad un correlatore, alla cui altra entrata perviene una versione ritardata dello stesso codice.

Alla uscita del correlatore si avrà uscita quando i due codici in entrata sono in sincronismo, ma questa condizione avviene unicamente quando il ritardo introdotto localmente è eguale al tempo di volo totale del segnale, ritardi strumentali inclusi.

Al solito, i ritardi strumentali, ionosferici, troposferici, di relatività vengono misurati e valutati e si perviene al calcolo della distanza.

I risultati sono interessanti: si ha, sempre per satellite geostazionario, una risoluzione dell'ordine di 3-4 cm ed una imprecisione che è limitata alla precisione con la quale si valutano o si misurano i vari ritardi. In pratica si resta attorno a 10-20 m, il limite è rappresentato dal valore del ritardo introdotto a bordo e dalla sua costanza nel tempo.

Sono in costruzione satelliti nei quali il ritardo del ripetitore di bordo potrà essere misurato a bordo periodicamente, inviandone poi il valore a terra.

L'interesse della misura di distanza di un satellite o sonda spaziale mediante un codice quasi aleatorio risiede nei seguenti fatti (i valori numerici si riferiscono ad una esperienza effettuata dal Politecnico di Torino sul satellite geostazionario Sirio 1):

— uso di potenza ridotta, (frazioni di watt, contro decine o centinaia di watt),
— uso di bande limitate (alcuni megahertz in luogo di decine di megahertz),
— uso di antenne di ridotte dimensioni (alcuni metri di diametro),
— risoluzioni dell'ordine del centimetro e possibilità di portare la precisione attorno o sotto un metro.

Ovviamente il principio di indeterminazione non è violato; infatti, usando:

— impulsi, si ha rilevante potenza e banda che sono impegnate unicamente durante il transito dell'impulso e l'atto di misurazione del tempo di volo si esaurisce in una frazione ridotta di tempo, stimabile attorno al microsecondo,
— i codici aleatori, il segnale viene emesso con continuità e la operazione di misurazione (la correlazione) è pure essa continua.

L'energia associata alle due misure è fondamentale la stessa, ma i vantaggi teorici e tecnici del secondo approccio sono ragguardevoli, soprattutto per la possibilità di sovrapporre il codice ad un altro segnale.

Altro rilevante interesse consiste in una proprietà statistica dei codici usati, la loro possibile ortogonalità che si manifesta con il fatto che due o più di siffatti codici hanno correlazione mutua idealmente nulla.

Questa proprietà consente che, attraverso il ripetitore del satellite, possano transitare allo stesso istante più codici, corrispondenti ad altrettante stazioni terrestri, ognuna delle quali può ricavare, *allo stesso momento dell'orbita del satelitte*, la distanza; è così possibile ricavare per multilaterazione l'orbita, mediante un procedimento puramente geometrico.

4.3. Impulsi luminosi

In luogo di impulsi radioelettrici, fig. 5, si usano impulsi luminosi.

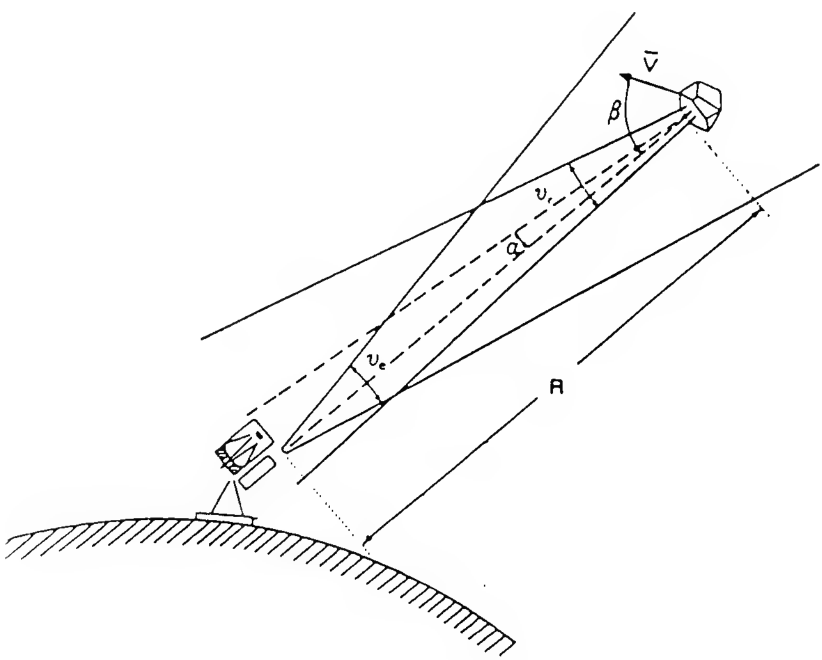


Fig. 5 - Misure di distanza di un satellite con tecniche laser. Parte dei fotoni generati da un laser ed emessi entro un determinato angolo solido, vengono raccolti da un retroriflettore e riemessi, entro un certo angolo solido, verso la stazione terrestre, ove una infima parte viene infine raccolta da un telescopio. Con satelliti geostazionari e specchi riceventi del diametro di un metro i fotoni contenuti nell'impulso di partenza, con taluni laser, ammontano 10E + 17 dei quali solo un centinaio sono raccolti a terra.

Numerosi i vantaggi:
— la portate, nel visivo, è attorno a 500 THz e, a queste frequenze, i ritardi ionosferici sono trascurabili,
— sono possibili larghezze di banda ragguardevoli, dell'ordine di molte centinaia di megahertz,
— l'impulso emesso da un laser può avere durate dell'ordine di 1 ns (l'impulso è lungo, nello spazio, 30 cm) e fronti di salita di centinaia di femtosecondi,

ma numerosi sono anche gli svantaggi:
— non esiste un «ripetitore» laser «imbarcabile», cioè un complesso ricevitore-trasmettitore laser, che possa sostenere i traumi di un lancio spaziale,
— di conseguenza la stazione a terra potrà ricevere solo la eco dei propri impulsi, da qui la necessità di usare rilevanti potenze in trasmissione o rilevanti specchi in ricezione, (vale l'equazione del radar, per la quale la ampiezza dell'eco scende con la quarta potenza della distanza),
— le dimensioni ridotte dei pannelli di retroriflettori che possono essere imbarcati sui satelliti, soprattutto perché un pannello piatto è causa di instabilità dell'orbita dovuta alle numerose forze che dipendono dalla superficie,
— la necessità che le condizioni atmosferiche siano adatte a questo tipo di misura,
— la necessità di disporre di numeroso personale specializzato.

Esistono alcune decine di stazioni laser che si dedicano a misure di distanza di adatti satelliti, prevalentemente per scopo geodetici e geofisici.

In Italia è attiva una stazione a Matera per Piano Spaziale Nazionale, un'altra è in corso di realizzazione a Cagliari ed a Torino, presso l'Aeritalia è stato costruito un apposito satellite, il LAGEOS 2 (Laser Geodetic Satellite) che verrà lanciato nel 1991.

Pur con queste limitazioni e problemi, la misura di distanza con impulsi è il metodo che consente al momento le massime precisioni, dell'ordine del centimetro per satelliti «bassi» cioè sino a 7-8000 Km e di 2-3 cm per l'orbita geostazionaria.

4.4. Misure di velocità relativa

Ogni portante a radiofrequenza emessa da un satellite perviene a Terra alterata dal noto effetto Doppler, fig. 6, dovuto al moto relativo satellite-stazione terrestre.

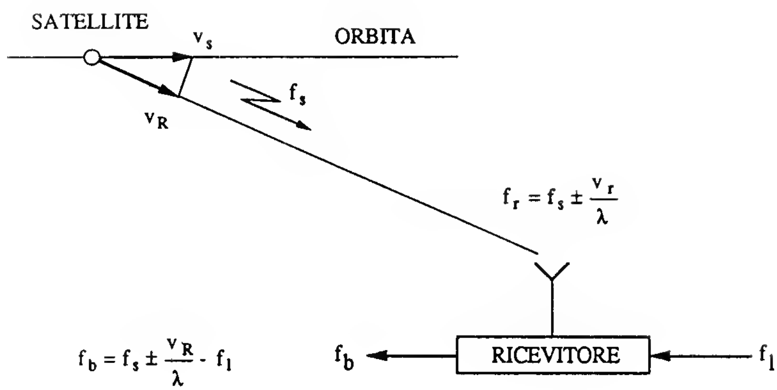


Fig. 6 - Principio del metodo Doppler o della misura della velocità relativa. Il segnale di frequenza f_s emesso dal satellite che ha una velocità relativa Vr rispetto a terra viene ricevuto, per effetto Doppler con la frequenza f_r. Nel ricevitore si fa battere un segnale locale di frequenza f_l con il segnale ricevuto, il segnale di battimento ottenuto, f_b, consente di ricavare la velocità relativa.

Sia il satellite posto su un orbita «bassa» e quindi animato di rilevante velocità relativa rispetto ad un punto fisso a Terra.

Il satellite emetta una frequenza campione e nota *fs* e quindi di lunghezza d'onda lambda nota, che perviene a Terra modificata; il segno *più* si applica se il satellite si sta avvicinando.

Entro il ricevitore, il segnale ricevuto viene mescolato con una frequenza campione e nota locale e si ottiene una frequenza di battimento, dalla cui misura si ricava, in modulo e segno, il valore della velocità relativa in funzione del tempo.

Il satellite irradia periodicamente un segnale di tempo e una equazione con la *stima* della propria orbita.

La stazione a Terra, posta in una posizione nota, calcola, in funzione del tempo, una serie di valori per la velocità relativa, usando la propria posizione e l'orbita ricevuta dal satellite.

Si ammetta che il transito del satellite duri venti minuti, durante i quali si individuino venti finestre temporali, ognuna della durata di venti secondi. Per ognuno di questi intervalli si effettua una *misura* della velocità relativa, il cui valore viene confrontato, intervallo per intervallo, con il corrispondente valore della stima. Ove le due serie temporali — quella delle *stime* e quella delle *misure* — non «coincidano», si altera l'equazione dell'orbita del satellite sino ad ottenere, entro certe tolleranze, concordanza tra le due serie temporali.

Le differenze di queste distanze individuano nello spaio degli iperboloidi, fig. 7, con i fuochi nelle posizioni istantanee del satellite.

Si opera per approssimazioni successive, con metodi di iterazione, sinché si arriva a determinare la posizione del satellite con errori dell'ordine di alcuni decimetri.

Solitamente, l'operazione viene invertita, nel senso che lo stesso tipo di misura — e quindi gli stessi strumenti e metodi di calcolo — vengono usati per determinare la posizione *incognita* di un punto a Terra. Le operazioni sono le stesse; l'unica differenza consiste nel fatto che la *stima* della velocità relativa viene calcolata assumendo un valore provvisorio per la posizione locale e ritenendo corretti i valori dell'orbita, quali irradiati dal satellite.

Il metodo Doppler, praticato in particolare con i satelliti americani per navigazione TRANSIT è stato il più diffuso nel periodo 1970-1985, consente di ricavare, con opportuni metodi sperimentali, la posizione di un punto sulla superficie terrestre con errori di 20 cm nelle tre coordinate, ed ha consentito numerose scoperte o miglioramenti nella conoscenza di grandezze di interesse per Geodesia, Geofisica ed Astronomia di posizione.

5. Sviluppi e nuove esperienze

La determinazione della distanza di un satellite al fine di determinare l'orbita, presenterà probabilmente nei prossimi anni quattro tipi di sviluppi:

— la diffusione dei codici quasi aleatori,
— la ricerca di sinergie tra metodi diversi,
— l'introduzione di laser con due colori, e
— l'introduzione di laser allo stato solido, in grado da essere lanciati.

Come esempio delle prime due tendenze, si consideri il problema della navigazione del satellite dell'ESA ERS1 (Earth Remote Sensing) che verrà lanciato tra alcuni anni. La navigazione è affidata ad un sistema chiamato PRARE (Precisione RAnge and range Rate Experiment), che consiste in misure di distanza tramite uno stesso codice quasialeatorio, sovrapposto a due portanti.

L'uso di due portanti, dato che il ritardo introdotto dalla ionosfera è funzione nota della frequenza, consente di ricavare le condizioni della ionosfera stessa, misurando la differenza tra i tempi di arrivo a Terra dei due codici, che partono simultaneamente dal satellite.

Le due portanti sono ricavate per sintesi da un campione di frequenza

e possono essere usate in un metodo doppler classico, sia separatamente, sia congiuntamente, per la compensazione dei ritardi troposferico. Una dclle portanti può essere rinviata al satellite e così si potrà avere una misura Doppler a due vie che presenta determinati vantaggi.

Il satellite è infine munito di un pannello di retroriflettori, per le misure laser.

Analogo sviluppo in progetto negli Stati Uniti, con il progetto ACRE (ACurate Range Experiment), che consiste nell'installare a bordo di uno speciale satellite della serie GPS (Global Positioning System), un pannello di retroriflettori e di un datatore a bordo dei tempi di arrivo degli impulsi luminosi da terra. In questo modo oltre a misure di distanza, saranno possibili misure di tempo.

Uno dei maggiori limiti alla misura della distanza è, come si è visto, la determinazione dei ritardi troposferici, usualmente effettuata mediante misure di pressione, temperatura ed umidità nei dintorni della stazione. La disponibilità, attualmente solo in laboratorio di ricerca, di laser emettenti contcmporaneamente su due o tre lunghezze d'onda, consentirebbe la correzione del ritardo dovuto all'indice di rifrazione dell'aria.

Infine, il giorno che fossero disponibili laser allo stato solido, affidabili, «robusti» al punto da poter essere lanciati e con la necessaria durata di servizio, sarebbe possibile sia avere un ripetitore laser sul satellite (restano comunque, ma possono essere risolti i problemi di puntamento), sia — e soprattutto — la interrogazione da parte del satellite di un gran numero di retroriflettori posti a Terra.

Infatti il costo di un retroriflettore di qualità adeguata è di quasi tre ordini di grandezza inferiore al costo di una stazione laser, la quale inoltre deve essere dotata di numeroso e scelto personale, mentre il retroriflettore, un oggetto passivo, una volta che sia protetto da vandalismi, può funzionare senza sorveglianza né manutenzione.

Inoltre tutti i dati verrebbero raccolti dal satellite, eliminando la attuale raccolta, spedizione e trattamento dei dati, operazioni di per se banali, ma onerose e non sempre sicure.

In questo modo il satellite potrebbe ricavare con continuità la propria orbita, interrogando contemporaneamente un numero sufficiente (almeno tre) di retroriflettori posti a Terra.

Ove questo laser imbarcato fosse a due colori, si avrebbe anche la correzione continua della rifrazione atmosferica.

Numerose sono le difficoltà tecnologiche, ma non sembra insperabili; l'obiettivo da raggiungere, nel giro di cinque-dieci anni, è una precisione di 5 mm sull'orbita di un opportuno satellite.

6. Alcuni risultati scientifici e tecnici consentiti dalla localizzazione precisa di satelliti

Numerosi, notevoli, ed in taluni casi con carattere di novità, i risultati che numerose discipline hanno potuto conseguire mediante la determinazione accurata di orbite; se ne darà poco più di un elenco.

Lo studio delle perturbazioni delle orbite consente di ricavare:

— l'andamento del geopotenziale attorno alla Terra; usando satelliti particolari od usuali su orbite di altezza diversa, il geopotenziale è oggi noto sino alla quarantesima armonica,

— come conseguenza, la distribuzione interna delle masse entro la Terra e la Luna,

— la superficie del geoide, con la scoperta e la misura di rilievi ed avvallamenti rispetto ad un ellissoide di riferimento,

— la scoperta di vulcani sottomarini nel Pacifico.

— nuovi elementi per determinare i tipi ed i valori delle forze non gravitazionali.

Lo studio delle orbite ha consentito misure nuove o di risolvere preesistenti problemi tecnici di misura per quanto riguarda:

— la determinazione, con 8-9 cifre del prodotto tra la costante di gravitazione universale e la massa della Terra,

— deriva dei continenti,

— polodia, cioè lo studio della migrazione dei Poli,

— variazione della lunghezza del giorno e quindi misure della velocità di rotazione della Terra,

— costituzione di reti topografiche, in maniera precisa e rapida in regioni non esplorate,

— la localizzazione di precisione di punti in mare aperto (piattaforme petrolifere, giunti di cavi sottomarini, ecc.),

— navigazione automatica di precisione marina, aerea, terrestre e spaziale,

— navigazione di sonde interplanetarie.

Come unico esempio di questi risultati, le figg. 7 e 8 riportano la forma del geoide a livello globale e la topografia dei mari attorno all'Italia, quali dedotte rispettivamente dallo studio di perturbazioni di orbite o da misure di altezza della superficie del mare effettuate da satellite la cui orbita era nota con precisioni, per quanto riguarda la componente radiale, dell'ordine del metro.

FIGURA 7

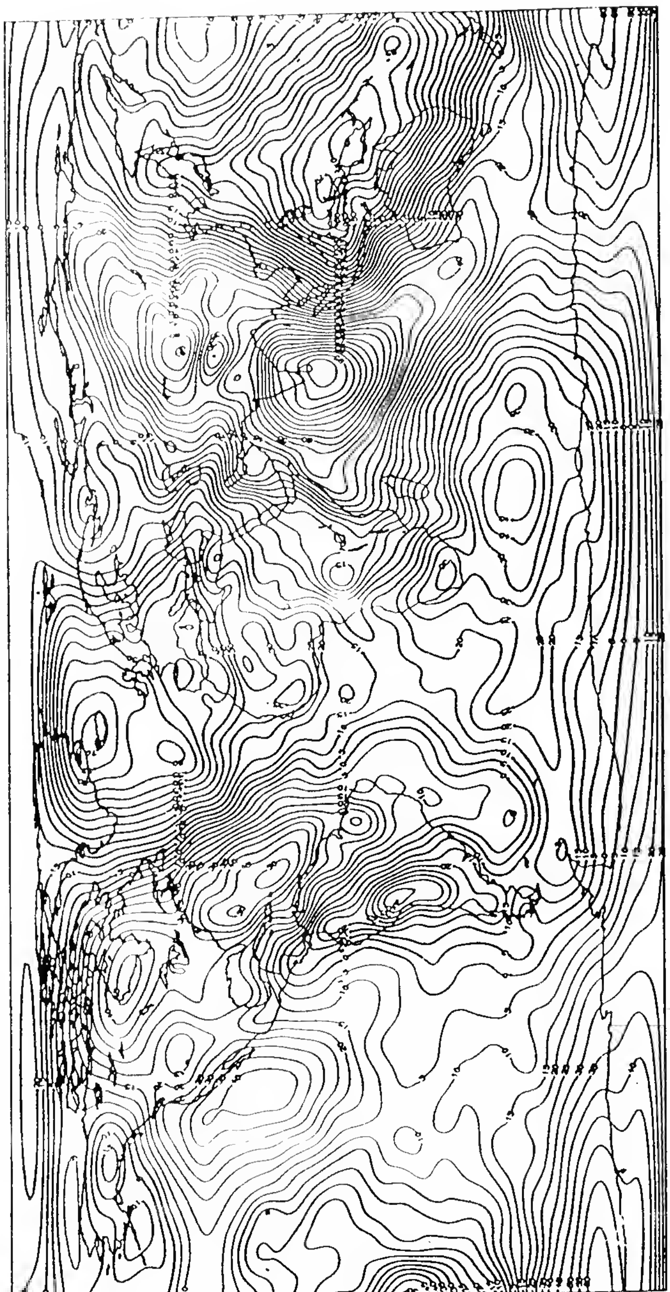


Figura 7 - Andamento su scala mondiale del geoide rispetto un ellissoide di riferimento; la equidistanza tra le linee di livello è di 5 m. Tra alcune depressioni, come quella a Sud della peninsola indiana ed alcuni «rilievi» si hanno differenze di quota del geoide che si avvicinano a duecento metri. La elaborazione è stata otenuta tramite lo studio di orbite misurate con l'effetto Doppler.

FIGURA 8

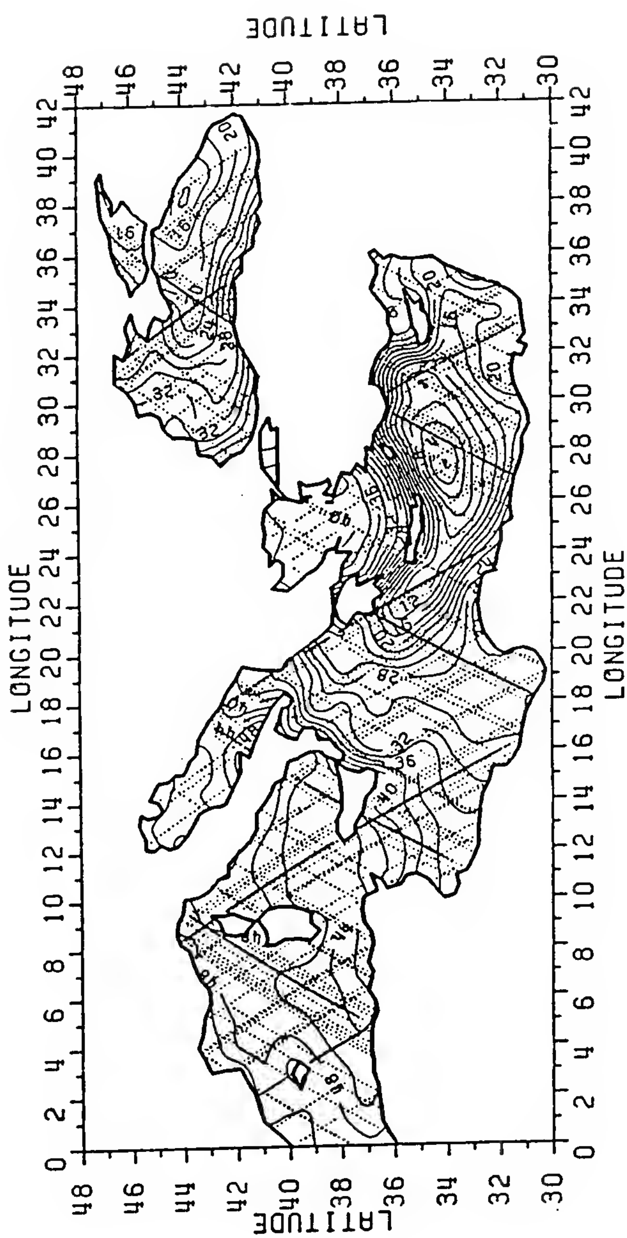


Figura 8 - Carta con le altezze medie del mare Mediterraneo e Mar Nero, rispetto un determinato ellissoide e ottenute mediante un radaraltimetro posto su un satellite la cui posizione era determinata con tecniche laser e Doppler. La equidistanza delle linee di livello è di due metri; dalla Sicilia alla Gracie, il Mediterraneo «scende» di quasi trenta metri.

Space Tethers Evolution of a new technology for space applications

Ernesto VALLERANI *

GIUSEPPE COLOMBO
SPACE TETHER PIONEER

A key figure in the development of the modern space tether concept was the Italian scientist Giuseppe Colombo. With an unrestrained imagination, he was able to foresee a wide range of potentially practical applications for space tethers and promoted them extensively despite early skepticism from the space community.

Colombo was a pioneer in modern problems of celestial mechanics. His understanding of the orbital dynamics of Solar System bodies and satellites paved the way for his many contributions to space research. Colombo's scientific interests were broad and extended well beyond tethers. He was, for example, an early proponent of a permanently manned space station and an active participant in the development of interplanetary and planetary space exploration missions.

Above all, he recognized the need for and strongly supported international cooperation in space. Colombo helped to guide Italy into joint projects with the United States, exemplified by the Tethered Satellite System currently under development, and consequently helped to stimulate the growth of the Italian space sector.

Colombo was born in Padua in 1920. He received a degree in mathematics in 1943 from the prestigious «Scuola Normale Superiore» at Pisa; served on the Faculty of Natural Sciences at the University of Padua; and in 1955 became a full professor of mechanics at the University of Catania. In 1962 he joined the Faculty of Engineering in Padua where he was given the Chair in Mechanical Vibrations, the first in this specialty in Italy, and later the Chair in Rockets and Spacecraft.

In 1961 Colombo went to the United States and began a long association with the Smithsonian Astrophysical Observatory (SAO) in Cambridge, Massachusetts. There he developed the first practical space tether application collaborating with, among others, Italian SAO scientist Mario Grossi. This work eventually led NASA to approve the joint Italy-U.S. Tethered Satellite System (TSS).

* Space Systems Group, Aeritalia, Turin.

Colombo's affiliation with Aeritalia began in the mid-1960s. He contributed extensively to studies that the company's industrial predecessors were conducting for early European satellites. Colombo continued this fruitful collaboration until his death in 1984, contributing his expertise to almost all of Aeritalia's space projects, in particular to initial development of the TSS and to studies of its future applications.

Colombo received a number of awards for his contributions to space research, the last of which was the NASA Gold Medal for Exceptional Scientific Achievement presented to him in 1983.

The spirit of Colombo's accomplishments was summarized by Roger Bonnet, Director of Scientific Programs for the European Space Agency at the February 20, 1985 conference held at the University of Padua in commemoration of Giuseppe Colombo. Said Bonnet: «... the real vocation of science is to break down the frontiers, discover the unknown, and is based on the rare and difficult convergence of imagination, rigorousness and development of new techniques».

SPACE TETHER CONCEPT

A simple space tether system consists of two masses connected by a wire. Because they are bound together, the masses are forced to orbit at the angular velocity of the system's center-of-gravity. Consequently, the speed of the upper mass is greater than it normally would be if the mass were untethered and orbiting at the same altitude. This mass experiences a greater centrifugal force than gravitational force. The opposite is true for the lower mass. Its speed is less than normal and it experiences a greater gravitational than centrifugal force.

The difference between the resulting gravitational and centrifugal forces on each mass creates tension in the tether that maintains the system in global equilibrium. This phenomenon is commonly called «gravity gradient».

There are a number of practical consequences of tethering. For example, gravity-gradient accelerations (often referred to as «artificial gravity») act on the system and increase in intensity along the tether length in the direction away from the system's center-of-gravity. This could be utilized for variable microgravity research. These gravity-gradient forces also tend to align the system along the local vertical which is the stable equilibrium configuration. Thus a tethered system could be used to help stabilize space structures during construction.

A variety of possible space applications for tethers are outlined in the following sections.

TETHERED SATELLITE SYSTEM

Space Shuttle engineering tests of the space tether concept are planned for the early 1990s using the Tethered Satellite System (TSS). This consists of an instrumented satellite; a thin, flexible tether up to 100 km long; a deployer attached to an enhanced Spacelab pallet in the shuttle cargo bay; and scientific experiments on the satellite as well as in the cargo bay.

The TSS is being developed jointly by Italy and the United States under an agreement between the Italian National Space Plan/National Research Council (PSN/CNR) and NASA.

The Aeritalia Space Systems Group is the Italian prime contractor responsible for the satellite, while Martin Marietta — Denver Aerospace is the American prime contractor responsible for the deployer and furnishing the tether.

NASA is providing the cargo bay pallet and experiment carrier. In addition, both NASA and PSN will provide the science payloads.

The TSS goal is to test the feasibility of deploying, controlling and retrieving a tethered satellite from the Space Shuttle, as well as to demonstrate the system's usefulness for scientific research.

Two TSS missions are currently planned: an electrodynamic mission using an electrically conducting tether plus satellite deployed spaceward; and an atmospheric mission with the tether and satellite deployed earthward. These will be discussed in more detail later.

Aeritalia is responsible for the TSS satellite design, development, construction, assembly, integration and qualification. The satellite is a sphere measuring 1.6 meters in diameter and weighing about 500 kg. It will house subsystems, such as propulsion, communications and thermal control, as well as scientific payloads. The current design also locates scientific instruments on one fixed and two deployable/retrievable booms extending from the satellite.

The following are the satellite's payload accomodation capabilities:

Payload mass	66 kg
Allowable area	3.4 m^2
Allowable volume	0.43 m^3
Thermal environment	$-10°C$ to $+50°C$
Electrical energy	2000 WH
Bit rate acquisition	10,250 KBPS max net
Telecommand bit rate	1,500 BTS max gross

Martin Marietta is responsible for development and construction of

the computer-controlled deployer. Its major components include a 12-meter collapsible boom and a reeling mechanism for the tether. The Cortland Cable Company is developing the tether for Martin Marietta. About 2 mm in diameter, it is made of duPont's Kevlar, a strong, lightweight, fibrous material. It is covered with Nomex for protection from the space environment. The electrically conductive tether for the first TSS mission will have an insulated copper-wire core.

During deployment, the satellite is unlatched from its support structure and is lifted clear of the Orbiter by the boom which extends vertically from the cargo bay. The satellite, with the use of small thrusters, is then released from the end of the boom as the tether is reeled out. At the end of the mission, the tether is reeled in.

TSS Electrodynamic Mission

The first TSS mission, planned for around 1991, will be dedicated to studies of the Earth's ionosphere and magnetic field.

A 20-km-long, electrically conductive tether will be deployed above the Shuttle (spaceward) to study the electrodynamic properties of the ionosphere as well as measure magnetic fields and plasma properties in this region. The mission duration will be about 36 hours. Of this, about 10 hours will be needed for tether and satellite deployment and about 7 hours for retrieval.

A variety of Italian and American scientifie experiments, with instruments located in both the satellite and cargo bay, will be conducted during this mission. These include studies of electrodynamic tether effects, plasma electrodynamics, magnetic fields, vehicle charging and potential, plasma coupling and dynamic noise.

This mission will also test the feasibility of generating electricity with the conductive tether. This could be an important source of power for future spacecraft and space stations. As the shuttle moves in Earth orbit, the tether will cut through the planet's magnetic field and interact with the ionospheric plasma. As a result, the electrically positive satellite will collect electrons from the ionosphere and these will be emitted back into space by an electron gun in the Shuttle cargo bay. This will cause a current to flow downward through the conductive tether and is expected to produce up to approximately 4 kW of power.

TSS Atmospheric mission

The second TSS mission, planned to follow the first by about two years, will be devoted to studies of the Earth's upper atmosphere.

The satellite will be deployed below the Shuttle (earthward) on a 100-km-long, non-conductive tether. In this configuration, the instrumented satellite will be lowered to an altitude of about 130 km above Earth. The aerothermodynamic properties of this region are still not well defined because of the difficulty of making direct measurements.

Atmospheric drag prevents satellites from staying in orbit long enough to make extensive measurements. Thus only small portions of this region have been probed with short-duration sounding rockets.

The satellite for the second TSS mission will be aerodynamically upgraded for maneuver control while being towed through the upper atmosphere. It will conduct a variety of scientific studies including direct measurement of magnetospheric, ionospheric and atmospheric coupled processes, thermospheric winds, neutral gas composition and temperature, ion composition, temperature and density, and ion drift velocity.

Data from this not-well-understood region is needed to improve global models of atmospheric chemistry and dynamics so that scientists can better evaluate such phenomena as the «hole» recently discovered in the atmosphere's protective ozone layer.

SPACE SHUTTLE TETHER DEMONSTRATION STUDIES

The space tether concept offers provocative new ways to utilize space that are radically different from traditional space systems.

The idea continues to gain attention and support from the spaee community as evidenced by a number of successful workshops that have been conducted, and are planned for the future, to explore the concept in depth and extract from it as many new applications as possible. The workshops are the results of a bilateral agreement between Italy's National Space Plan (PSN) and NASA. The first was held in 1983 in Williamsburg, Virginia, followed by others in 1985 (Venice, Italy) and in 1986 (Washington, D.C.). A fourth is scheduled for October 1987 in Venice.

Out of these formal exchanges came a multitude of ideas, some of which today may seem more science fantasy than science. Consequently, PSN and NASA faced the difficult challenge of selecting ones for further definition by industry; ones that hold the most promise for future application to the International Space Station and other large autonomous spacc structures.

Aeritalia is studying several of these ideas. They include a tethered space elevator and a tether-based re-entry system for returning vehicles to Earth. The latter is particularly important now that is has been determined that the download mass for the proposed Space Station is severely limited. A simple tethered system could help ease the situation.

The initial concepts selected for further definition could be validated by Space Shuttle demonstrations. Unfortunately, with the Challenger disaster, not only is the first tether demonstration (Tethered Satellite System) delayed, but so too these additional demonstrations and consequently the development pace for this new technology.

The following section briefly highlights the 10 space tether concepts initially selected by PSN and NASA for further definition and validation. They are, however, not the only ones that could or should be studied. In fact, new ideas continue to come forward and the initial selection of studies no doubt will be modified accordingly.

Tethered elevator test flight

A module or «space elevator» capable of crawling in a controlled manner along a deployed tether appears to have a number of potential Space Station applications. These include: a variable microgravity laboratory; transportation unit between two tethered bodies; system center-of-gravity management; tether inspection and repair device; and carrier for re-entry probes.

The design for a scaled-down space elevator, whose performance could be demonstrated with the Tethered Satellite System, is being developed by the Aeritalia Space Systems Group. The elevator's characteristics include:

— $0.65 \times 0.65 \times 1.05$ m in dimensions,
— 70 kg mass,
— 2 m/s maximum velocity,
— 100 W power consumption,
— one-axis attitude control,
— S-band communications,
— friction drive mechanism,
— grapple fixture for Shuttle Remote Manipulator System,
— front slot for positioning elevator on tether.

Power/Thrust generator

A tethered system could be used to generate DC electrical power to supply on-board loads or, used in the reverse mode, to obtain an electromagnetic propulsive thurst to boost a spacecraft's orbit.

A Space Shuttle demonstration of the power/thrust generator would use an electrically conductive tether, 10-20 km long, at the end of which is attached a satellite. The purpose is to demonstrate the feasibility of producing up to 1 MW of power or up to 200 N of thrust at about 90 percent efficiency.

Electrical power is generated as the tether moves through the Earth's magnetic field. The attached satellite, acting as a plasma contactor, collects electrons from the ionosphere which are then emitted into space by an electron gun in the Shuttle cargo bay. This causes an electric current to flow downward through the tether. The first Tethered Satellite System mission is designed to test the feasibility of this concept.

Generating electric power with a tether produces a decelerating force that lowers the spacecraft's orbit. Conversely, reversing the power generation process could be used to generate thrust for boosting the spacecraft's orbit. This is done by feeding current into the tether from an on-board power supply. The electric current running upward through the tether creates an electromagnetic propulsive force causing the spacecraft's altitude to increase.

Tethered «Wind Tunnel»

The Space Shuttle can be used to demonstrate a tethered system's potential as an open, continuous wind tunnel facility.

Attached to the end of a tether, an instrumented aerodynamic model can be deployed below the Shuttle and towed through the upper atmosphere at altitudes around 100-150 km. Thus the model can be exposed over a long period of time to conditions that are nearly impossible to reproduce in ground-based wind tunnels, including regimes of low Reynolds number and large Mach number.

A Space Shuttle-based wind tunnel demonstration would use a 100-km-long tether. Instruments in the aerodynamic model would either record or telemetry data on such factors as heat transfer and drag coefficients, air flow and turbulence. The data can be used to evaluate designs for new-generation aerospace vehicles.

Furthermore, a tethered wind tunnel facility could yield extensive ex-

perimental data about the aerothermodynamics of this part of the upper atmosphere whieh currently can only be probed directly with short-duration sounding rockets.

Tether-initiated re-entry system

A tethered system combined with an elevator carrier could be used to return materials to Earth or dispose of them by atmospheric burn-up without the use of rockets.

A capsule containing experimental samples, data, etc. would be attached to the elevator carrier. This, in turn, would move the capsule to the appropriate point along the tether to finely control its re-entry trajectory before release. The capsule's landing would be controlled by parachute. If the capsule contains material to be destroyed, its trajectory would be controlled so that it would burn up in the Earth's atmosphere on re-entry rather than land.

A tether-initiated re-entry system could be applicable to the International Space Station. A Space Shuttle demonstration of the method would use hardware from the Tethered Satellite System (deployer, satellite); a 30-km-long tether; and a capsule mass of 80 kg. In addition, it would require the development of an elevator carrier, engineering instrumentation, a parachute system, and a thermal protection system.

Tethered orbital transfer system

Tether dynamics make it possible to deploy a payload into a higher orbit from a spacecraft in low-earth orbit using momentum transfer.

A Space Shuttle demonstration of a tethered orbital transfer system would use a variable length tether released spaceward. At its end would be a launching platform holding a satellite to be deployed.

With the tether extended, the Shuttle is below the system's center-of-gravity; momentum from the spacecraft is transferred to the platform/satellite which is above CG. Thus, when the satellite is released, its higher-than-normal momentum boosts it to a higher orbit, even into geostationary orbit under certain conditions. At the same time, the Shuttle's less-than-normal momentum will cause its orbit to lower, assisting its return to Earth.

Such a tethered orbital transfer system results in fuel savings for both satellite deployment and Space Shuttle re-entry.

Tethered pointing system

Active control of a movable attachment point on a tethered body represents a new method of attitude control. This could be used for high precision pointing of a science and applications platform tethered to the Space Station.

The performance of such a tethered pointing system could be demonstrated with the Tethered Satellite System using an improved satellite. This would feature a movable attachment point as well as a new attitude measurement and control system, and engineering instrumentation.

Tens of arcseconds pointing performance is expected. Space Shuttle demonstration goals include: measurement of attitude dynamics with and without attitude control; measurement of attitude stabilization in response to induced dynamical disturbances; and measurement of displacement mechanism and control system performance.

The tethered pointing system is one of several new tether application coneepts being developed at Aeritalia.

Tether-aided assembly

Because of the gravity gradient that results from tethering, a tethered system is self-orienting along the local vertical. This feature could be used during space construction activities to maintain the correct orientation of structures.

During the assembly phase, space structures may tend to be unstable and may require an attitude control system to keep them oriented properly. Passive control may be possible by attaching one or more tethers with end masses for ballast. The advantages include: system simplicity, relatively low costs and reusability. Disadvantages include the small probability that the tether will be severed.

Requirements for a tether-aided assembly system include:
— stabilization by either single or double tether arrangement,
— simple one-way deployer with expendable tether,
— 1-2 km tether length,
— up to 2,000 kg tether end mass,
— small tether control capability may be required.

Long space antenna

A tethered system, using an electromagnetic tether, could be the basis for an orbiting long antenna to generate ULF/ELF/VLF waves for worldwide communications.

For a Space Shuttle demonstration of this concept, the spacecraft would be connected to an insulated conducting tether that is deployed spaceward and ended with a mass that serves as an electron collector.

A giant loop antenna in space is created by the interaction of the tether with the magnetic field and ionosphere in the following manner: electrons from the ionosphere spiral along magnetic field lines and are attracted by the electron collector at the end of the tether. The electrons move down the tether and are shot back into space from an electron gun in the Shuttle cargo bay. They then spiral along magnetic field lines back into the ionosphere closing the current loop.

This giant space antenna generates electromagnetic waves into which messages can be incorporated by modulation, for example, by turning the electron gun on and off at the desired frequency. Theoretically, these waves may be spread quickly over the globe by a process called ducting to provide instant worldwide communications.

Frequencies of 30-30 Hz (ULF), 3-300 Hz (ELF) and about 3000 Hz (VLF) could be produced by a Space Shuttle demonstration using a 20-to-100-km-long tether with a 10 A tether current. This current level would inject about 1 W of power by night and 0.1 W by day into the transmission line.

Tethered multiprobe for atmospheric studies

Conducting atmospheric studies at different levels with good time correlation of measurements is possible by stringing several instrumented probes along a tether deployed earthward from a spacecraft.

This one-dimensional constellation appears to be particularly valuable for low-altitude studies and when data has to be collected simultaneously at different atmospheric regions.

A Space Shuttle demonstration of the multiprobe system would utilize the TTS deployer and satellite, and include up to five probes attached to a tether. Requirements include an operational sequence of deployment and retrieval, and possibly a crawling mechanism.

Get-Away tether experiment

A small free-flying tether system could be released from the Space Shuttle to perform a variety of basic tether experiments. These include demonstrating electromagnetic power/thrust generation; measuring tether radar cross section; assessing impact hazards to the tether; and carrying out radio propagation experiments.

This so-called «get-away» tether experiment would consist of a 20-kg upper mass and a 50-kg lower mass connected by a 1-km-long tether.

Initially, the tether would be coiled and the two masses joined at deployment. Separation, with the tether played out by a miniature reeling mechanism, would take about one-half hour. At 24 hours after deployment the system would be fully separated and stabilized.

After about 60 to 90 days of experimentation, the mission would be ended possibly by severing the tether to obtain additional information about its survivability in the space environment.

FUTURE SPACE STATION TETHER APPLICATIONS

The full exploitation of the innovative space tether concept can only be realized in terms of application to a large space project such as the proposed International Space Station.

This is the goal as scientists and engineers work out the details for Space Shuttle demonstrations of the initial group of concepts selected by PSN and NASA for further definition and validation.

The advantages of tethers to a space station hadn't escaped the imagination of space tether pioneer Giuseppe Colombo. He recognized early on that tethers could revolutionize the architecture of a future space station. Even before such a structure was formalized by the United States, the Italian scientist had developed specific applications. At the time, some seemed much too fantastic to be practical, but now they appear to be valid applications or, at least, could be evolved into realistic applications to a space station.

One of Colombo's futuristic ideas is that of a space port to serve the space station. It would be constructed of tethered spent external tanks from Space Shuttles. Others include a variable microgravity laboratory, a tethered science platform, a tether-based deployer, and power generation via tether.

The intense level of activity currently surrounding space tethers un-

derscores the vitality of the concept. It is hoped that its enthusiastic support from the space community can transform its futuristic/idealistic image into one of a sound technology. Only then will space tethers emerge as a practical new tool for future use in space.

The following section briefly highlights five possible Space Station applications for tethers. As with all things that deal with the future, it is impossible to predict which ones will ultimately prove practical.

Space port

A space port that would serve the Space Station could be constructed out of tethers and spent external tanks from the Space Shuttle.

The Shuttle external tanks are currently disposable, but they could be recycled by latching them together to form two rafts separated by tethers. The resulting structure could serve as a space port for the Shuttle and other spacecraft. In addition, it would offer a «space anchor» for other tether applications, for example, an astrophysics platform, tethered payload launcher, earth observation platform and zero-energy elevators.

Such a space port would require tethers 10-to-20 km long, reconfiguration of the external tanks, and assembly/build-up operations.

Elevator released re-entry system

The space elevator used as a transportation device could be part of a tethered return/disposal system for the Space Station.

Materials to be sent back to Earth or to be disposed of would be put into a re-entry capsule. This would be carried by the space elevator to the appropriate point along the tether for the desired trajectory: either one that permits landing on Earth, or one that causes burn-up in the atmosphere on re-entry.

A permanently deployed Space Station re-entry system would include a tether up to 70-km long and a 7-ton space elevator.

This sytem could support a 3-ton re-entry capsule. For returning materials to Earth, the capsule would require a thermal protection system as well as a parachute system for landing. This would not be needed for a disposal capsule.

The advantages of this system for the Space Station include quick re-entry capability as well as fuel and cost savings.

Two other tether-based re-entry systems are currently being studied by Aeritalia. One would use an intermittently deployed tether with the re-entry capsule attached to an end bus. The other would be a disposable system where the capsule would be released by cutting the tether.

Tethered space elevator

The tethered space elevator could play an essential role in expanding the capabilities of the Space Station. Deployed on a tether that is attached at one end to the Station and at the other to a large platform, the elevator could perform a number of functions.

The most promising is that of a variable microgravity laboratory to take advantage of gravity-gradient accelerations along the tether length. These increase in strength with distance from the system's center-of-gravity and are perceived as variable levels of microgravity. For a tethered system in low-earth orbit, the attainable microgravity level at CG may be about 10^{-8} g and increases by about 4×10^{-4} g/km.

A space elevator would be able to position experiments at desired microgravity levels. Of particular interest to the international science community is the elevator's unique capability to provide a time-varying gravity profile.

Other important applications of the tethered space elevator are transportation between Station and platform, and overall center-of-mass management.

The advantages of a space elevator include easy deployment on the tether and ready accessibility to payloads. Such a system would require docking capabilities at the platform and Station ends. It would have a 5-ton mass with a 3-ton payload accommodation capacity, and a maximum cruising speed of 1 m/s.

If needed, a secondary cable could provide a power and data link. This would be 12 kW of power and 40 Mbps data rate for the microgravity laboratory application. For the transportation application, 2 kW power and 64 Kbps data rate would be provided by on-board subsystems.

Astrophysical and Earth survey facility

Space Station science payloads that require highly accurate pointing could be accommodated on a tethered pointing platform.

Pointing is achieved by active control of a movable attachment point

on the platform permitting 3 degrees of freedom in movement and tens of arcseconds pointing performance.

The 10-ton platform could accommodate a 5-ton payload mass. A tether power line would supply 15 kW of power and an optical fiber link would carry data at a rate of 20 Mbps.

The facility would allow periodic payload reconfiguration and provisions for different emergency conditions.

Multiple tether applications for long-term Space Station capability enhancement

Multiple tethered systems could be used for the long-term enhancement of Space Station capabilities. Many of the tethered systems described in this report are applicable, by they represent only a small part of what is possible with tethers.

Possible uses for an upward deployed tethered system include:
- Orbital Transfer Vehicle (OTV) hangar/refueling facility that is intermittently deployed;
- second tether reeled out from hangar for momentum transfer of OTV;
- propellant depot for spacecraft refueling;
- retrieval of OTV hangar with a second tether deployed.

For a downward deployed tethered system, the following are possible multiple Space Station applications:
- large tethered platform (50 tons) for science and technology applications;
- tethered space elevator for platform servicing, variable microgravity laboratory, re-entry system, tether inspection and repair.

Utilization of libration-point orbits

David W. DUNHAM*

Abstract. *This paper presents an overview of proposed astrodynamical uses of orbits near libration points. The first artificial use was in 1978, when a scientific spacecraft called the third International Sun-Earth Explorer (ISEE-3) was placed into a «halo orbit» around the Sun-Earth L1 libration point. ISEE-3 also used lunar swingbys to explore the Earth's geomagnetic tail, and then left the Earth-Moon system to become the first spacecraft to encounter a comet, Giacobini-Zinner, in 1985.*

Plans are described for using trajectories similar to ISEE-3's for the many spacecraft of the International Solar-Terrestrial Physics program to make a more comprehensive study of the Sun and the magnetosphere during the next decade.

Introduction

In 1772, the French mathematician J. Lagrange demonstrated that there are five positions of equilibrium in a rotating two-body gravity field [Reference 1]. Three of these «libration points» are situated on a line joining the two attracting bodies, and the other two form equilateral triangles with these bodies. More information about libration points can be found in Szebehely's comprehensive book on the restricted problem of three bodies, especially in Chapter 5 [Reference 2]. As shown in Figure 1, a total of seven libration points are located in the Earth's neighborhood and consequently may have practical astronautical applications. Five of them are members of the Earth-Moon system and two belong to the Sun-Earth system. In the reference frame used here, the Sun-Earth line is fixed and the Earth-Moon configuration rotates around the Earth.

The relative stability of motion near the L4 points of the Earth-Moon system have attracted the most popular attention as proposed sites for space colonies [Reference 3]. Possible periodic orbits about L4 are shown in Figure 2. Finding orbits about L4 or L5 with long-term stability is not trivial due to strong solar perturbations [Reference 4].

* Computer Sciences Corporation Lanham-Seabrook, Maryland 20706, U.S.A.

The linearized equations of motion in a rotating coordinate system centered on the Sun-Earth L1 point are shown in the upper-left corner of Figure 3. If the exponential terms in the equations of motion have zero coefficients, only the periodic terms shown in the lower-right part of Figure 3 remain. The frequencies of the motion in x and y are the same, so the in-plane trajectory is an ellipse centered on the libration point, as shown in the upper left panel of Figure 4. The frequency of the motion in z is slightly different, resulting in the lissajous patterns for the out-of-plane motion shown in the lower panels of Figure 4.

Although motion near the collinear points is unstable, very little propulsion is needed to keep a spacecraft at or near one of them for an extended period of time. The first serious proposals to send a spacecraft to the neighborhood of a libration point have involved the collinear points. Prof. Colombo pioneered in work on control of a spacecraft near collinear points of the Earth-Moon system, showing [Reference 5] that:

(1) To the first order, the collinear points are exact solutions of a restricted four-body problem (Sun, Earth, Moon, and spacecraft).
(2) It is possible to use a simple linear feedback control for position stabilization.
(3) It is possible to obtain the acceleration needed for position and attitude stabilization of a libration-point spacecraft by simply varying the magnitude and direction of a solar sail.

In this same work, Colombo was also the first to suggest that varying the length of a eable conneeting two tethered spacecraft might be used to produce stable motion near a collinear point. A variety of stationkeeping techniques for spacecraft in the vicinity of a collinear libration point is discussed in Reference 6.

If the amplitudes of a trajectory about a collinear point are increased, higher-order terms in the motion change the frequencies slightly. It is then possible to find initial conditions where the inplane and out-of-plane frequencies are equal [Referenees 7, 8, 9]. Halo orbits may play an important role in future lunar operations. In a concept originally proposed twenty years ago [References 10, 11], it was shown that a data-relay satellite located in a halo orbit around the Earth-Moon L2 point could provide an uninterrupted communications link between the Earth and the far side of the Moon (see Figure 5). This capability will be essential for a manned base on the Moon's farside. In the more immediate future, a farside communications satellite could facilitate the navigation and control of unmanned lunar vehicles in scientifically interesting areas such as the Tsiolkovsky crater.

The ISEE-3 halo orbit mission

The first use of a libration-point orbit by man involved the L1 point of the Sun-Earth sysrem. Several years ago, it was recognized that this point is an ideal location to continuously monitor the interplanetary environment upstream from the Earth [Reference 12]. A suitably instrumented spacecraft placed in the vicinity of this point can provide data on the solar wind about an hour before it reaches the Earth's magnetosphere. However, the spacecraft cannot be stationed too close to the L1 point because the Sun is dircctly behind this point when viewed from the Earth. This alignment is a problem because the intense solar noise background will severely degrade downlink telemetry. Figure 6 shows a typical lissajous orbit as seen from the Earth. For the first two circuits, there is no problem. But after that, the trajectory repeatedly crosses the shaded solar interference exclusion zone. To avoid this zone, a halo orbit is utilized [Rcference 8]. This orbit has a period of approximately six months and passes slightly above and below the ecliptic plane, as shown in Figures 7 and 8.

The halo orbit is inherently unstable, leading some critics to doubt the practicality of a libration-point mission; see Figure 9.

Stationkeeping maneuvers are required to keep the spacecraft close to the halo path. Fortunately, the timescales are relatively long and studies showed that stationkeeping maneuvers are only needed about once every three months. Their associated delta-V costs are quite small, only about 10 m/sec per year.

On August 12, 1978, the third International Sun-Earth Explorer (ISEE-3) spacecraft was launched toward the Sun-Earth L1 point. The spinning spacecraft is illustrated in Figure 10. Its primary goal was to continuously monitor the solar wind before it interacted with the Earth's magnetosphere, which was being studied by the ISEE-1 and ISEE-2 satellites in lower highly elliptical orbits. ISEE-3 has a full complement of experiments for studying plasmas and the electromagnetic environment, listed in Table 1. The current status of the experiments is shown in the last column.

Two possible transfer trajectories to the L1 point are shown in Figure 11. There is no scientific advantage for the fast transfer; useful data about the solar wind can be obtained as soon as the spacecraft crosses the magnetospheric bow shock less than two days after launch for either trajectory. Earlier studies for transfers to the lunar L2 point showed that the total delta-V requirement could be substantially reduced by adding maneuvers at strategic points on the trajectory (see Figure 12).

TITLE	PRINCIPAL INVESTIGATOR	AFFILIATION	EXPERIMENT STATUS
SOLAR WIND PLASMA	BAME	LOS ALAMOS NATIONAL LAB	ELECTRONS ONLY (ION PORTION FAILED)
PLASMA COMPOSITION	OGILVIE	GSFC	OPERATIONAL
MAGNETOMETER	SMITH	JPL	OPERATIONAL
PLASMA WAVES	SCARF	TRW SYSTEMS	OPERATIONAL
ENERGETIC PROTONS	HYNDS	IMPERIAL COLLEGE, LONDON	OPERATIONAL
RADIO WAVES	STEINBERG	PARIS OBSERVATORY MEUDON	OPERATIONAL
X-RAYS, LOW ENERGY ELECTRONS	ANDERSON	UCB	X-RAYS AND E_e >200 keV (LOW ENERGY ELECTRON PORTION FAILED)
LOW ENERGY COSMIC RAYS	HOVESTADT	MPI	PARTIAL FAILURE (ULEZEO)
MEDIUM ENERGY COSMIC RAYS	VON ROSENVINGE	GSFC	OPERATIONAL
HIGH ENERGY COSMIC RAYS	STONE	CIT	PARTIAL FAILURE (ISOTOPE PORTION)
HIGH ENERGY COSMIC RAYS	HECKMAN	UCB/LBL	PARTIAL FAILURE (DRIFT CHAMBER)
COSMIC RAY ELECTRONS	MEYER	UNIVERSITY OF CHICAGO	OPERATIONAL
GAMMA RAY BURSTS	TEEGARDEN	GSFC	PARTIAL FAILURE (PHA MEMORY)

Table 1 - ISEE-3 Experiments.

	RADIAL (IN-PLANE) COMPONENT			AXIAL (OUT-OF-PLANE) COMPONENT (m/sec)
MANEUVER	PLANNED (m/sec)	REALIZED (m/sec)	EXECUTION ERROR	
MIDCOURSE #1	16.98	17.74	4.5%	0.11
MIDCOURSE #2	25.02	24.29	2.9%	0.54
HALO INSERTION	11.73	11.73	–	2.30
STATIONKEEPING #1	2.01	1.94	3.5%	–
STATIONKEEPING #2	1.52	1.58	4.0%	–

Table 2 - Summary of ISEE-3's Early Delta-V Maneuvers.

Even greater savings can be made by inserting the spacecraft into a halo orbit about the libration point, rather than travelling to the point itself, which needs to be avoided in any case by ISEE-3 to avoid the solar radio interference zone.

ISEE-3's actual transfer trajectory is shown in Figure 13. One hundred days after launch, on November 20, 1978, ISEE-3 was inserted into a halo orbit around the L1 point, thus becoming the first libration-point satellite. Three orbit maneuvers with a total delta-V expenditure of 54 m/sec were needed to achieve the desired mission orbit. During the halo-orbit phase (November 1978 to June 1982), fifteen stationkeeping maneuvers with a delta-V cost of 30 m/sec were required for orbit maintenance. The first two stationkeeping maneuvers are indicated on Figure 13. The radial and axial components of these maneuvers, along with those performed in the transfer phase and the error in the execution of each maneuver determined from post-maneuver orbit determination, are listed in Table 2. Further details of the ISEE-3 flight performance from 1978 to 1982 can be found in References 13 and 14.

Double lunar-swingby orbits

Long-term investigation of the distant regions of the geomagnetic tail is a major goal of space plasma science. A spacecraft located in a halo orbit around the Sun-Earth L2 point could provide data at distances between 220 and 250 RE from the Earth. However, a trajectory that allows repeated longitudinal scans of the magnetotail between 60 and 250 RE is preferred. Cross-sectional coverage at various downstream distances is also favored. Conventional Earth orbits will not satisfy these specifications because the apsidal line of an elliptical Earth orbit is essentially fixed in inertial space and appears to rotate about one degree per day with respect to the Sun-Earth line, as shown in Figure 14. Therefore, to obtain the desired coverage, it will be necessary to control the rotation of the apsidal line to maintain the apogee segment in the tail region, since the geomagnetic-tail axis is always aligned within a few degrees of the Sun-Earth line; see figure 15. The required orbital rotation could be achieved by employing propulsive maneuvers, but the delta-V cost would be about 400 m/sec per month! This reality led to the development of a technique that uses a series of lunar gravity-assist maneuvers for orbital control, another imaginative use of a special class of orbits in the three-body problem.

The basic procedure is depicted in Figure 16. Assume that a space

craft is initially located near the apogee of the smaller orbit at A1. After its next perigee passage, the timing of its motion is planned so that the Moon is at the «x» above «S1» when the spacecraft crosses the lunar orbit on its outbound leg. The trailing edge lunar swingby maneuver at S1 will then rotate the line of apsides counterclockwise by an amount of 1/2 delta-omega and will also raise the apogee distance to A2. A leading-edge lunar swingby at S2 (symmetric with S1), after the Moon has completed one full orbit plus the S1-S2 arc, will return the spacecraft to its original orbit, but will rotate the line of apsides counterclockwise again by an amount 1/2 delta-omega. This sequence of orbit pairs could be repeated indefinitely; one more cycle is shown in Figure 16. The orbit is periodic in a geocentric reference frame rotating at the lunar rate, as shown in Figure 17. Notice that the Moon does not occult the trajectory from an observer on the Earth.

Broucke first found orbits like this in a survey of periodic orbits in the restricted problem of three bodies using the Moon/Earth mass ratio [Reference 12]. But he did not realize its practical significance, discussed below.

Both the small and large orbits can be modified to vary the size of delta-omega. If it is about 60 degrees, as shown in Figure 16, the mean rate of rotation of the line of apsides will be about one degree per day, since the repeat cycle time is two months. The rate can be adjusted to match the mean rate of the Earth's motion around the Sun (called the solar rate below) shown in Figures 14 and 15. This has been done for the orbit shown in Figures 16 and 17. The same orbit is shown in the upper section of Figure 18, in a reference frame rotating at the solar rate. Since the geomagnetic tail points in the anti-Sun direction, it remains relatively fixed in this reference system, pointing to the right near the Sun-Earth dashed line. Since the orbit is closed in the rotating frames of both figures 17 and 18, it has the unusual characteristic of being doubly periodic.

The time spent in the outer loop between S1 and S2 is just over one month, as noted above. For this reason, we call this a one-month class double-lunar swingby orbit. The lunar swingby distance (perilune radius) in kilometers, and the apogee and perigee radii expressed in Earth radii (Re) of 6378.16 Km, are given in the upper section of Figure 18. But other solar-rate double lunar-swingby orbits can be designed by decreasing the lunar swingby distanee. This causes a larger increase in the orbital energy and semi-major axis of the large orbit, causing the spacecraft to achieve a higher apogee at A2.

The orbit shown in the middle section of Figure 18 spends a little more than two months in the outer loop for a total cycle time of 3 months.

The velocity of this two-month class orbit is less than the rotational velocity of the coordinate system at the outer-loop apogee, which produces a twist in the orbit. With an even smaller swingby distance (but still a comfortable 8 lunar radii above the Moon's surface), a three-month outer looop results shown in the bottom section of Figure 18. The twist is expanded due to the lower velocities in the distant parts of the orbit, which extends nearly to the L2 Sun-Earth libration point. The three classes of orbits illlustrated in Figure 6 represent just a few of the many solutions that can be formed with the double lunar-swingby technique. Additional solutions can be obtained by increasing the time interval as well as the number of orbital loops in the inner trajectory segment. Four- and five-month outer loops are also possible. Details of these solutions are given in References 13 and 14.

The periodic orbits in Figures 16, 17, and 18 have been generated with a simplified patched-conic dynamical model. When a more realistic model that inclused the effects of solar perturbations and the Moon's orbital eccentricity is used, the symmetrical shapes are distorted and the apogee distances are changed. Nevertheless, the average apsidal rotation can be kept at the required Sun-synchronous rate.

ISEE-3's geotail excursion

The first flight application of the double lunar-swingby concept came in 1983 when it was used by the ISEE-3 spacecraft [Reference 15]. Figures 19 and 20 show the ISEE-3 trajectory from June 1982 to September 1983. These are rotating ecliptic-plane views, with fixed Sun-Earth line. The geomagnetic tail is shaded, with its mean aberration angle of 4 degrees below the anti-Sun direction shown. A delta-V maneuver of only 4.5 m/sec was used to initiate the transfer from the halo orbit to the geotail. Except for small navigational corrections, this maneuver was sufficient to complete the flight path shown in Figure 19. However, a large out-of-plane maneuver was needed at Al to change the inclination of the ISEE-3 trajectory and set up the lunar encounter at S1. As shown in Figure 20, swingby maneuvers at S1 and S2 produced a five-month geotail traverse that extended out to an apogee distance of 237 RE. The propulsive maneuver on June 1, 1983 was needed to satisfy targeting and phasing constraints for a gravity-assist maneuver on December 22, 1983 that ejected ISEE-3 from the Earth-Moon system, as shown in Figure 21.

Double lunar-swingby orbits are highly unstable. An error of a few kilometers in one lunar swingby usually results in an error of hundreds

of kilometers at the next swingby, and a complete miss at the one after that. But like the halo orbit, the time scale of the motion is rather long. The several weeks between lunar swingbys give time to calculate and perform needed small correction maneuvers, and determine the orbit accurately between maneuvers. The history of correction maneuvers performed to correct ISEE3's third and fourth lunar swingbys, following the large orbit adjust maneuver on June 1, 1983, is shown in Figure 22. The desired lunar swingby distance, and the actual (observed) distance derived from orbit determination following each maneuver, are listed at the bottom of the Figure. The trajectories of ISEE-3 following the June 1st, July 1st, and August 12th maneuvers, assuming that no further corrections were performed, are shown in the next three figures. In Figure 23, the third swingby (S3) was on the wrong side of the Moon.

After the July 1st correction shown in Figure 24, the swingby is on the correct side of the Moon, and an inner-loop orbit is formed, but S4 is missed. Following the August 12th maneuver, the inner loop is corrected so that S4 occurs, along with a trajectory resembling the two month outer loop to S5 shown in Figure 21. The later maneuvers refined the trajectory even more, so that the error in the critical fifth swingby was only 2 kilometers. ISEE-3's double-lunar swingby trajectory is also shown in the inertial equatorial-plane view of Figure 26. The value of the orbit is not apparent unless it is portrayed in the solar-rotating views of the previous several figures.

In addition to their effectiveness for orbital control, lunar-swingby maneuvers can be used to supply the energy increase needed to escape from the Earth-Moon system and enter a heliocentric orbit [References 16 and 17]. As early as January 1959, a lunar flyby was used to catapult the Soviet space probe Luna-1 into a solar orbit. However, using the Moon to send a spacecraft toward a specific target in heliocentric orbit is a more sophisticated task. At the lunar swingby, the aim point in the impact plane, the relative velocity vector bertween the spacecraft and the Moon, and the timing of the lunar encounter are precisely related. This type of lunar-swingby maneuver was performed for the first time on December 22, 1983, when ISEE-3 was targeted for an encounter with Comet Giacobini-Zinner.

ISEE-3's lunar flyby on December 22, 1983 just grazed the Moon's surface as shown in Figure 27. For nearly half an hour, ISEE-3 was in the lunar shadow, the first eclipse experienced by the spacecraft since launch. This caused some concern, since the batteries had failed while in the halo orbit a couple of years before. Hydrazine in the fuel lines might freeze, rupturing them and rendering ISEE-3 uncontrollable. The

spacecraft heaters were turned on high as it approached the Moon. A few minutes after emerging from the shadow, ISEE-3 turned itself on as it received power from its solar cells, and resumed transmitting, much to the relief of ground controllers. Another view of ISEE-3's trajectory around the Moon is shown in Figure 28, and an artist's view of the fly-by is depicted in Figure 29.

A total of five lunar swingbys and fifteen propulsive maneuvers were needed to carry out the transfer from the halo orbit to the escape trajectorey. The four planned orbit-adjust maneuvers accounted for 72 m/sec of the total delta-V cost of 77 m/sec. Maneuver execution errors and orbit determination uncertainties were responsible for the remaining 5 m/sec. Immediately following the successful S5 escape lunar swingby maneuver, NASA announced that it had decided to change the name of ISEE-3 to the International Cometary Explorer (ICE).

Even though ICE passed extremely close to the Moon, it could only achieve a hyperbolic excess velocity of 1.67 km/sec (C3 = 2.8 km2/sec2). This value is close to the practical limit for a single escape maneuver. However, by using a special double lunar gravity-assist technique [Reference 18], it is possible to reach a higher energy level. The use of the double gravity-assist technique was considered in formulating a contingency plan for the ISEE-3/ICE departure. This alternative escape trajectorey is shown in Figure 30. Notice that, in this example, a hyperbolic excess velocity of 2.12 km/sec (C3 = 4.5 km^2/sec^2) is produced.

Ice's heliocentric orbit and comet encounter

ICE became the first spacecraft to encounter a comet on September 11th, 1985. The orbits of ICE, the Earth, and Comet Giacobini-Zinner (GZ) are shown in the isometric view of Figure 31. An ecliptic-plane view is depicted in Figure 32. ICE's historical passage through GZ's tail is shown in Figure 33. The target point had to be carefully selected. If ICE passed too close to the nucleus, it might be destroyed by high-speed impacts with cometary dust. If the distance from the nucleus was too great, ICE might miss the tail, which wags in an unpredictable manner in the solar wind. To make matters worse, the comet's orbit changes due to irregular outgassing as the comet nucleus heats up near perihelion. The orbit changed a few thousand kilometers two weeks before the encounter, requiring a maneuver of over 2 m/sec three days beforehand [Reference 19]. Fortunately, there was very little dust, and ICE suffered no noticeable damage. Yet, the comet's tail was four times larger than ex-

pected, enabling ICE to gather a rich harvest of information about GZ's complex plasma environment [Reference 20]. ICE passed through the neutral sheet at the center of GZ's tail. Before the encounter, ICE was given only one chance in ten of accomplishing this feat. An artist's rendition of the encounter is shown in Figure 34.

ICE's heliocentric trajectory, in a rotating frame with the Sun-Earth line fixed, is shown in Figure 35. ICE's heliocentric period is about 12 days shorter than the Earth's, so its mean motion is greater.

Consequently, it slowly drifts away and ahead of the Earth. But near apogee, its velocity decreases, becoming less than the Earth's due to the larger eccentricity of its orbit. That is the reason for the annual loops in the trajectory. Besides studying comet GZ, ICE also made valuable up-stream measurements of the solar wind affecting Halley's Comet in October 1985 and March 1986. Energetic particles from Halley itself may have been detected during the latter approach.

In the rotating frame, the trajectory continues to loop away from the Earth. Eventually, it drifts all the way around, and will pass close to the Earth again in the year 2014. Manuevers were performed in 1986 to return the spacecraft to the Earth's vicinity, depicted in Figure 36. Specifically, ICE has been targeted for a close lunar swingby on August 10, 2014. This can capture the spacecraft back into an Earth trajectory, shown in Figure 37, that is nearly a mirror-image of the escape trajectory in Figure 21. The miss distance at the lunar swingby is now calculated to be about 21,000 km, but this can be decreased little now due to the inaccuracies of ICE's control system and uncertainty in the orbit (especially from solar radiation pressure on the slowly-precessing satellite) over nearly 30 years [Reference 21]. A year or two before encounter, ICE can be more accurately targeted for the lunar swingby. After the capture, ICE's orbit can be further lowered with propulsive and lunar swingby maneuvers. The spacecraft then might be taken aboard an orbit maneuvering vehicle (OMV) and returned to Earth. A detailed orbit operations handbook has been written to help mission controllers in the 21st century maneuver ICE [Reference 22]. If a new comet comes close to ICE's path, the spacecraft's remaining delta-V capacity of about 150 m/sec might be used to encounter it, which would change the orbit and preclude a return to Earth. Unfortunately, ICE does not pass close enougth to any known comets to arrange an encounter before 2014.

International solar terrestrial physics program

The International Solar Terrestrial Physics (ISTP) program will involve several spacecraft to be built and launched by the National Aeronautics ad Space Administration (NASA), the European Space Agency (ESA), and the Japanese Institute of Space and Astronautical Science (ISAS). The spacecraft will be launched during the early 1990's to study in detail the geomagnetic environment, the Sun, and their complex interactions. The ISTP program will be coordinated with spacecraft that have already been, or will be, launched by other agencies, and the Soviet Interkosmos agency may formally participate in ISTP. Figure 38 is a schematic diagram (now slightly out-of-date) showing the ISTP spacecraft, as well as spacecraft making related measurements (Ulysses, UARS, Spacelab/SOT, and the space station).

The launch dates and planned orbits for the ISTP spacecraft are shown in Table 3. The Equator spacecraft was originally envisioned by NASA, but had to be dropped due to budget constraints. Scheduling problems and other concerns prevented a proposed «Equatorial Science Phase» for one of the Cluster spacecraft, to be launched early by NASA and maneuvered using lunar swingbys to joint the other 3 spacecraft after ESA launched them. A Soviet spacecraft may serve the role of Equator, but with different parameters than those listed in the table.

Table 3 - ISTP orbits

Spacecraft	**Built by**	**Lauch Date**	**Lauch by**	**Perigee × Apogee Radius/ Inclination/Period**	**Remarks**
SOHO	ESA	3/95	NASA	250 Re/Sun-Earth L1 Halo/6 months	
Cluster	ESA	12/95	ESA	3×20 Re/90 deg./55 hours	4 Spacecraft
Equator	?	9/93	?	2×12 Re/0 deg./26 hours	see text
Geotail	ISAS	7/92	NASA	8×80-250 Re/Lunar/2-6 months	1st year
				8×30/16 deg./117 hours	2nd year
				8×80-250 Re/Lunar/2-6 months	3rd year
Wind	NASA	12/92	NASA	8×80-250 Re/Lunar/2-6 months	1st year
				250 RE/small L1 «halo»/6 months	2-3 years
Polar	NASA	6/93	NASA	2×9 Re/90 deg./18 hours	apogee north

The success of the ISEE-3 libration-point mission has spawned a number of follow-on proposals. As part of its contribution to ISTP, ESA is planning to station a solar observatory called SOHO in an ISEE-3 type halo orbit in 1995 [Reference 23]. A libration-point orbit around the Sun-Earth L2 point has been mentioned as a possible location for a Soviet Prognoz mission in the early 1990s. Finally, it is expected that, during the 2nd phase of its mission, the Wind spacecraft (to be provided by NASA) will occupy a Sun-Earth L1 small-amplitude halo orbit where it will serve as a replacement for ISEE-3 discussed below.

Three projects scheduled for implementation in the 1990s are planning to use double lunar-swingby trajectory profiles. One of them is a Japanese mission called MUSES-A that will be launched sometime in 1990 [Reference 24]. A second Japanese spacecraft, Geotail, will follow in 1992.

The third example, NASA's Wind spacecraft (mentioned in the previous section), will reside in a sunward double lunar-swingby orbit for about a year before it is subsequently placed into a Sun-Earth L1 halo orbit. An important goal for Wind will be to monitor the solar wind upstream from the Earth, but from a trajectory that stays close to the Sun-Earth line for longer periods of tine than the broad halo orbit of ISEE-3. The solar wind is not uniform, so data representing the wind impinging on the Earth's magnetosphere are best obtained close to the Sun-Earth line. The two-month-class double lunar-swingby trajectory shown in Figure 39 will accomplish this task.

A trajectory whose departure begins near the Earth with velocity just under that for escape is shown in Figure 40. With a higher velocity, the spacecraft would move past the Sun-Earth L1 point and escape towards the Sun. With less velocity, it would enter a highly elliptical orbit with the apogee pointing towards the Sun, but not reaching L1. The trajectory shown is a fine balance between these conditions; the spacecraft enters an orbit similar to ISEE-3's halo orbit with no transfer or insertion maneuvers needed, except for small navigational corrections. Direct transfers like those shown always result in orbits with a large Y amplitude, passing far from the Sun-Earth line. But by using a lunar swingby, it is possible to go into a halo-like orbit with a much smaller Y amplitude, as shown in Figure 41. Wind's last lunar swingby will be similar to this, to affect a transfer from the double lunar-swingby phase to a small-amplitude «halo» orbit about L1 with almost no delta-V. By varying the lunar swingby time and distance, it should be possible to reach even smaller orbits about L1. The «blocky» appearance of the trajectories in these figures is due to the motion of the Moon. If the orbits

are plotted in a solar-rotating frame centered at the Earth-Moon barycenter, rather than geocentric, the bends straighten out and the trajectories become quite smooth.

Concluding remarks

It has been shown that libration-point orbits and trajectories using lunar gravity-assist maneuvers can provide a large degree of orbital flexibility for spacecraft in the Sun-Earth-Moon system. The practicality of libration-point orbits and lunar-swingby maneuvers has been fully demonstrated by the flight of ISEE-3/ICE. Additional flight experience with these trajectory techniques will be gained from at least five more missions scheduled for launch in the 1990s.

Acknowledgements

I am deeply indebted to Robert Farquhar at NASA's Goddard Space Flight Center. He had most of the basic ideas discussed above, as can be inferred by scanning the references. He directed the work of myself and of my industrious co-workers, at both Computer Sciences Corporation and at Goddard, to prove the practicality of these techniques using realistic computer models. Farquhar also fostered contacts with key scientists and administrators that were crucial for obtaining approval for the ISEE-3 mission and its spectacular extensions. For this paper, he provided much useful information and many of the figures. Finally, we are all indebted to Prof. Giuseppe Colombo for his pioneering work and brilliant ideas about libration-point orbits, and about many other areas of celestial mechanics and astrodynamics.

REFERENCES

[1] LAGRANGE J., 1873 - *Oeuvres* (M.J.A. Serret, ed.), Vol. 6, Gauthier-Villars, Paris. Lagrange's main article on the three-body problem was dated 1772, but not originally published until 1777 by Panekoucke, Paris; see page 304 of Reference 2.

[2] SZEBEHELY V., 1967 - *Theory of Orbits - The Restricted Problem of Three Bodies*, Academic Press, New York and London.

[3] O'NEILL G., 1974 - *Physics Today 27*, No. 9, 32.

[4] SCHUTZ B.E., 1977 - Orbital mechanics of space colonies at L4 and L5 of the Earth-Moon system, *AIAA Paper* 77-73.

[5] COLOMBO G., 1961 - The stabilization of an artificial satellite at the inferior conjunction point of the Earth-Moon system, *Smithsonian Astrophysical Observatory Special Report* No. 80.

[6] FARQUHAR R.W., 1968 - The control and use of libration-point satellites, *Stanford University Report* SUDAAR-350 (reprinted as NASA TR R-346, 1970).

[7] FARQUHAR R.W. & KAMEL A.A., 1973 - Quasiperiodic orbits about the translunar libration point, *Celestial Mechanics 7*, 458-473.

[8] FARQUHAR R.W., MUHONEN D.P. & RICHARDSON D.L., 1977 - Mission design for a halo orbiter of the Earth, *Journal of Spacecraft and Rockets 14*, 170-177.

[9] BREAKWELL J.V. & BROWN J.V., 1979 - The «Halo» family of 3-dimensional periodic orbits in the Earth-Moon restricted 3-body problem, *Celestial Mechanics 20*, 389-404.

[10] FARQUHAR R.W., 1966 - Stationkeeping in the vicinity of collinear libration points with an application to a lunar communications problem, *AAS Preprint* 66-132.

[11] FARQUHAR R.W., 1967 - Lunar communications with libration-point satellites, *Journal of Spacecraft and Rockets 4*, 1383-1384.

[12] FARQUHAR R.W., 1969 - Future missions for libration-point satellites, *Astronautics and Aeronautics 7*, 52-56.

[13] FARQUHAR R.W. & DUNHAM D.W., 1981 - A new trajectory concept for exploring the Earth's geomagnetic tail, *Journal of Guidance and Control 4*, 192-196.

[14] DUNHAM D.W. & DAVIS S.A., 1980 - Catalog of double lunar-swingby orbits for exploring the Earth's geomagnetic tail, *Computer Sciences Corporation Report* TM-80/6322.

[15] FARQUHAR R., MUHONEN D. & CHURCH L.C., 1985 - Trajectories and orbital maneuvers for the ISEE-3/ICE comet mission, *Journal of the Astronautical Sciences 33*, 235-254.

[16] CAMARENA V., 1970 - «Lunar spring-board» effect, *La Recherche Aerospatiale 2*, 63-68.

[17] CAMARENA V. & MARCHAL C., 1971 - Complement to the study of the lunar spring-board effect, *La Recherche Aerospatiale 2*, 119-121.

[18] ROSS D.J., 1980 - Material capture by double lunar gravity assist, *AIAA Paper* 80-1673.

[19] ROBERTS C.E., 1985 - International Cometary Explorer (ICE) maneuvers 92 Through 99, *Computer Sciences Corporation Report* TM-85/6117.

[20] VON ROSENVINGE T., BRANDT J.C. and FARQUHAR R.W., 1986 - The International Cometary Explorer mission to the comet Giacobini-Zinner, *Science 232*, 353-356. This is the lead article of several in this issue that give the first scientific results from the ICE flyby of GZ.

[21] ROBERTS C.E., 1987 - International Cometary Explorer (ICE) maneuvers 100 through 110, *Computer Sciences Corporation Report* TM-87/6011.

[22] ROBERTS C.E., 1987 - Handbook for International Cometary Explorer (ICE) orbit operations, *Computer Sciences Corporation Report* TM-87/6039.

[23] ESA SCI (85) 2 1985, SOHO: Report on the Phase A study.

[24] UESUGI K., HAYASHI T. & MATSUO H., 1986 - MUSES-A double lunar swingby mission, *IAF Paper* 86-294.

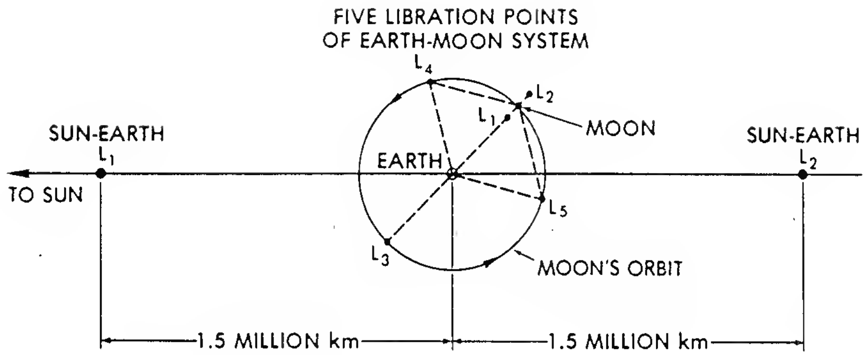


Figure 1 - Libration Points in the Vicinity of the Earth.

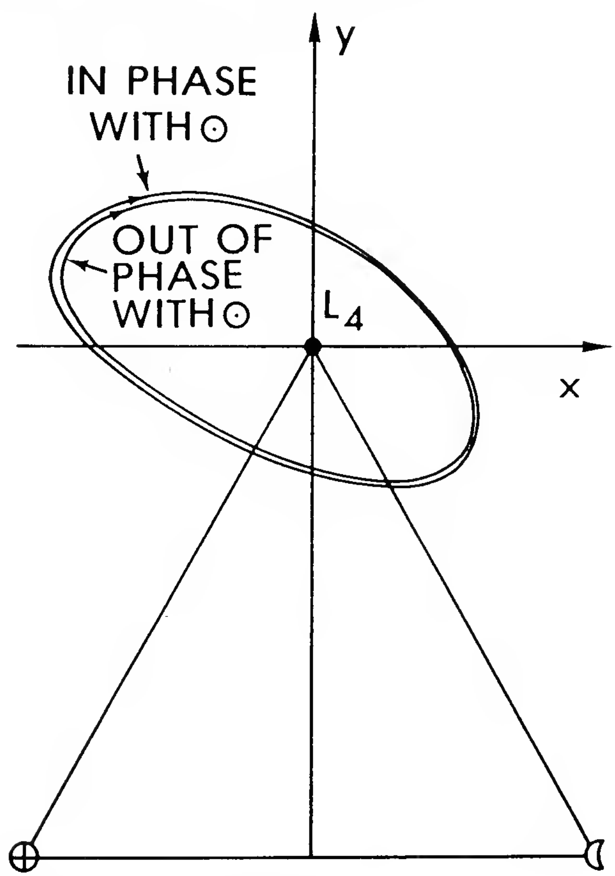


Figure 2 - One-Month Stable Periodic Orbits.

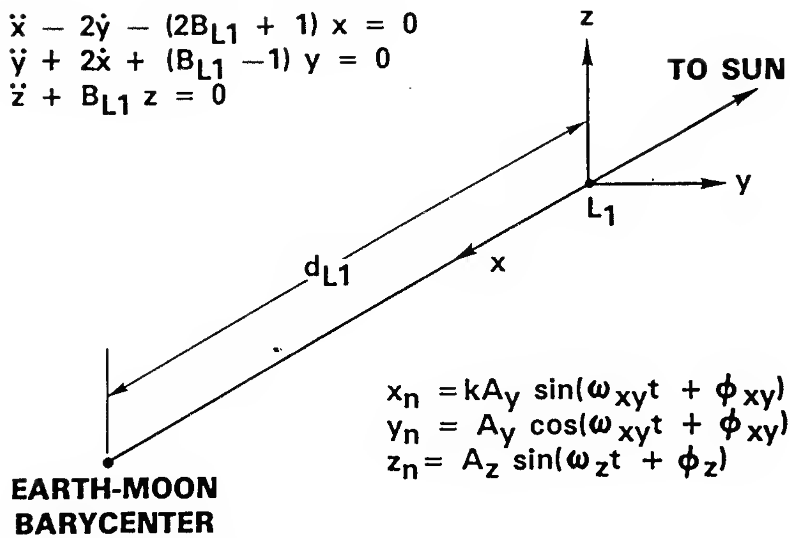


Figure 3 - Equations of Motion near the Sun-Earth L1 Libration Point.

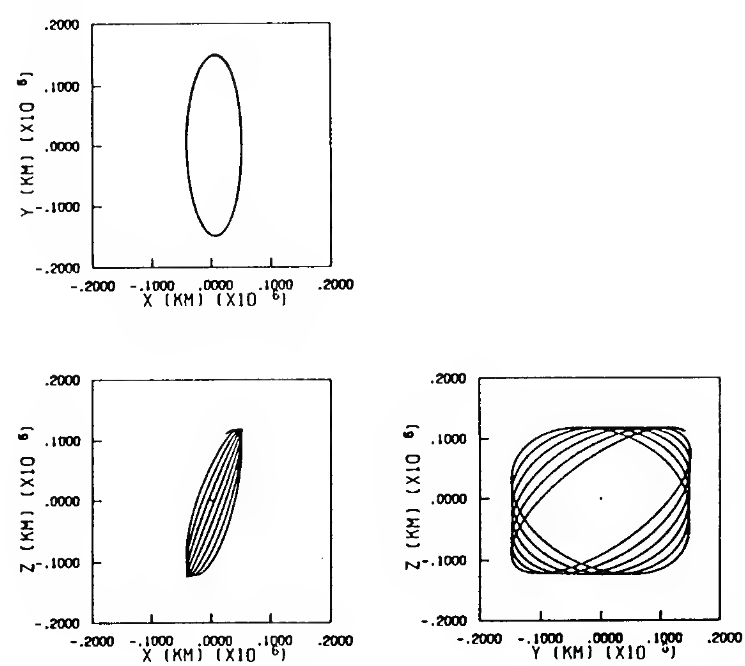


Figure 4 - Three-year Lissajous Path near the L1 Libration Point, Amplitudes Az = 120,00 km and Ay = 150,000 km.

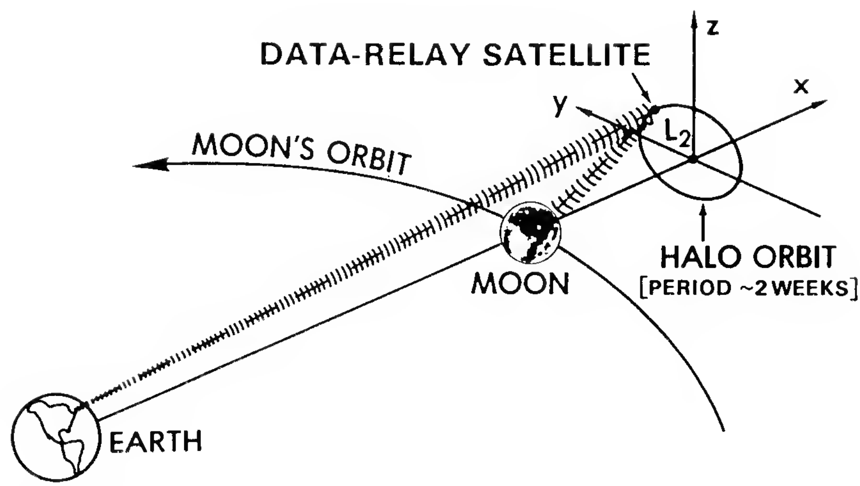


Figure 5 - Lunar Farside Communications Link.

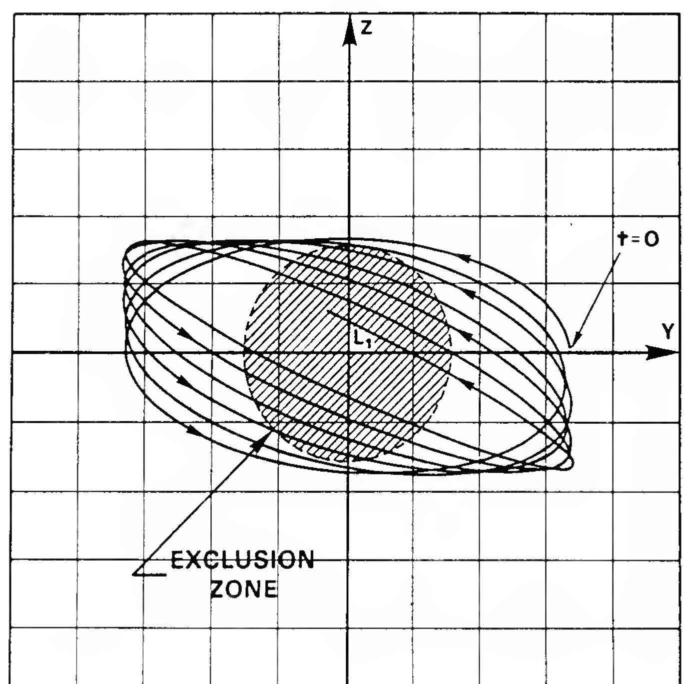


Figure 6 - Lissajous Path crossing Solar Radio Interference Zone.

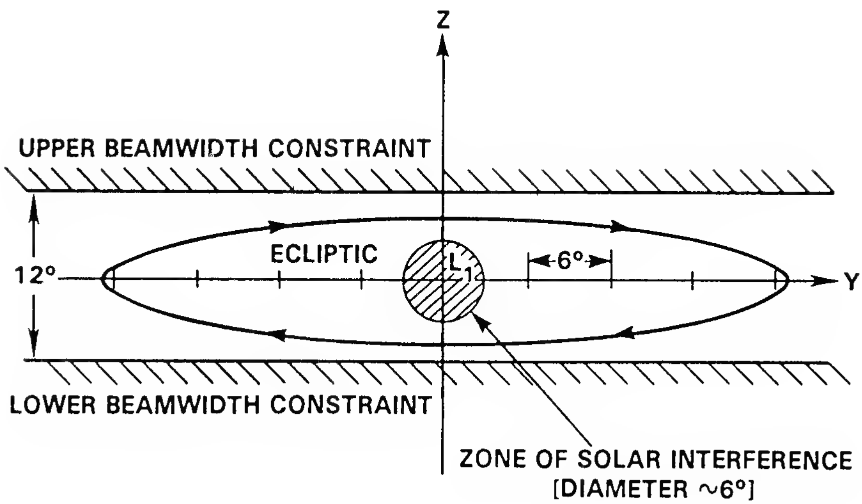


Figure 7 - Halo Orbit as seen from Earth.

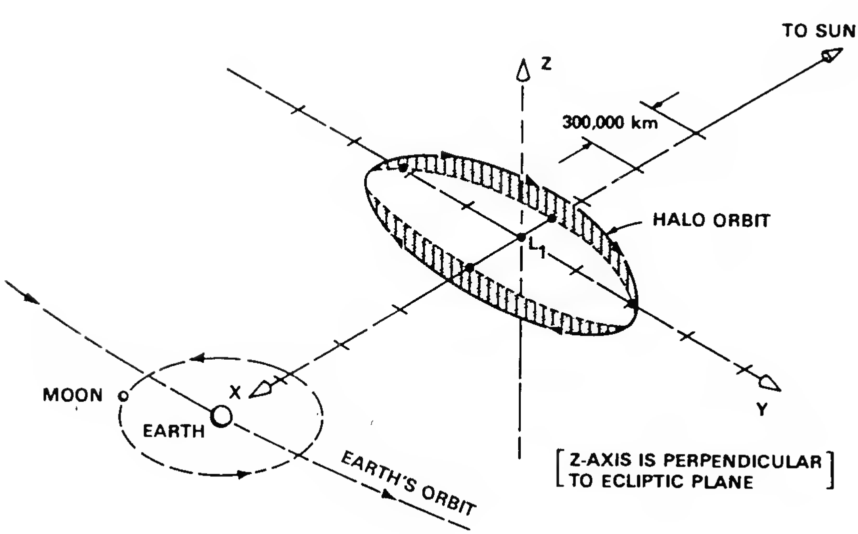


Figure 8 - Halo Orbit Around the Sun-Earth L1 Libration Point.

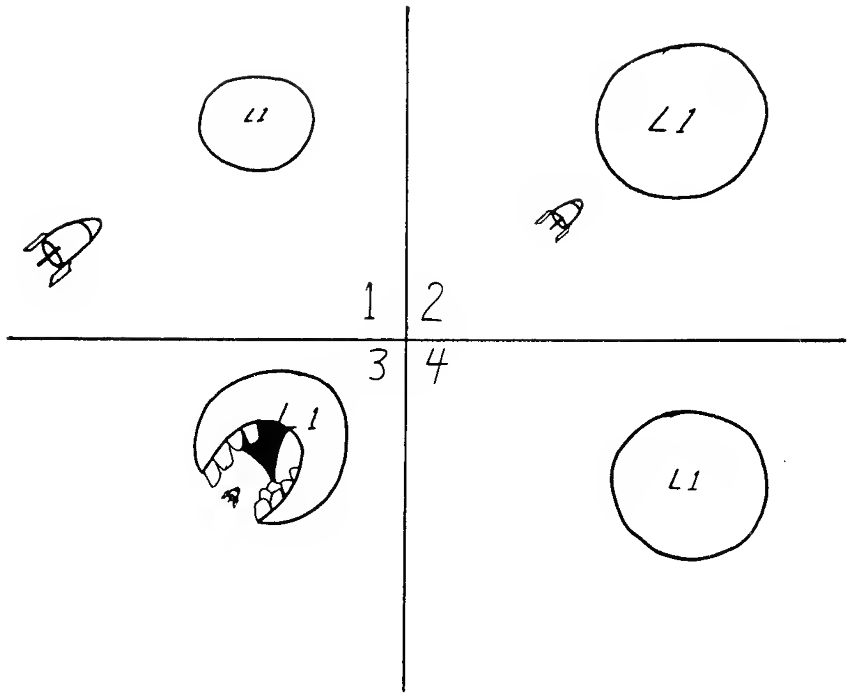


Figure 9 - Fear of Travelling near a Libration Point.

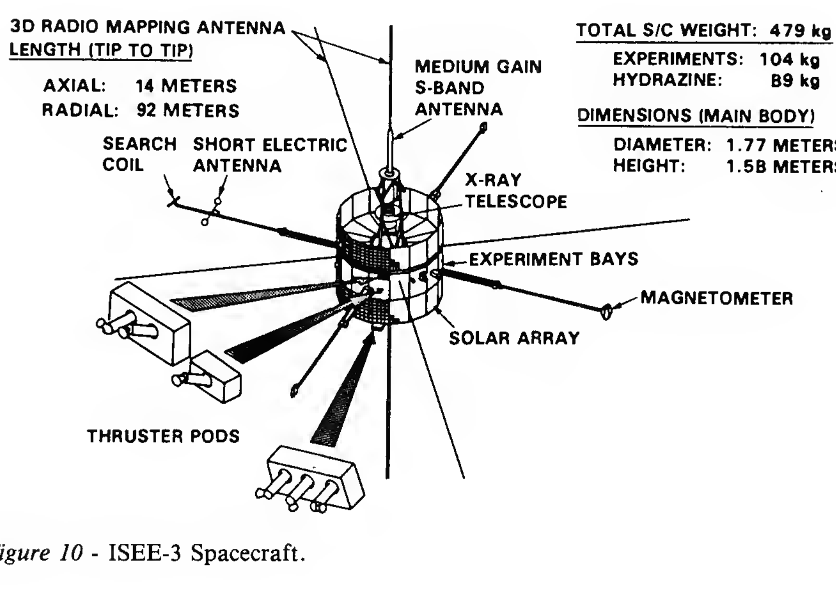


Figure 10 - ISEE-3 Spacecraft.

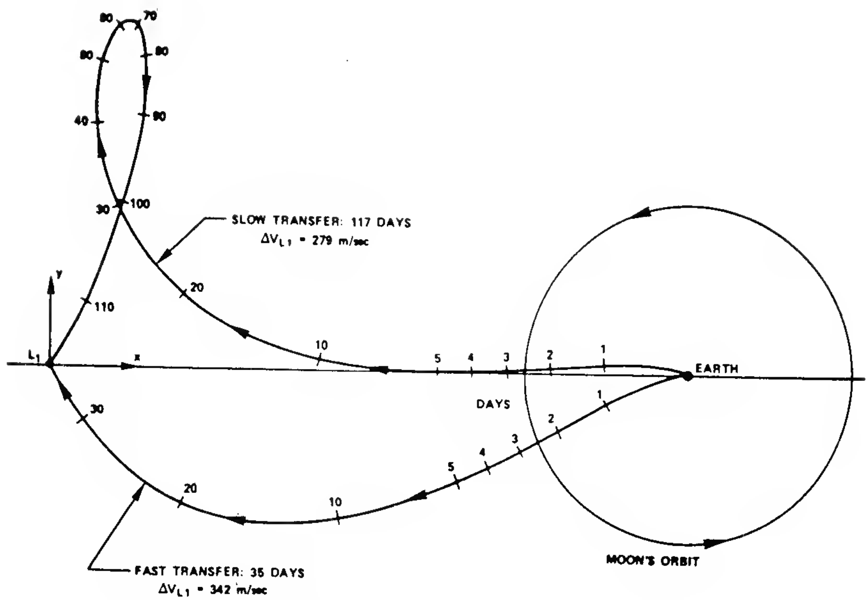


Figure 11 - Possible Transfers to L1.

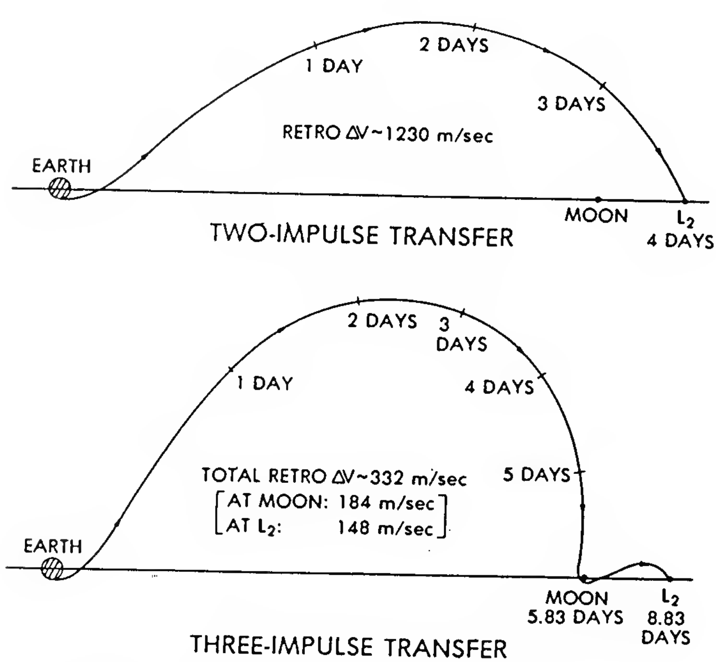


Figure 12 - Trajectories to Vicinity of Earth-Moon L2 Point.

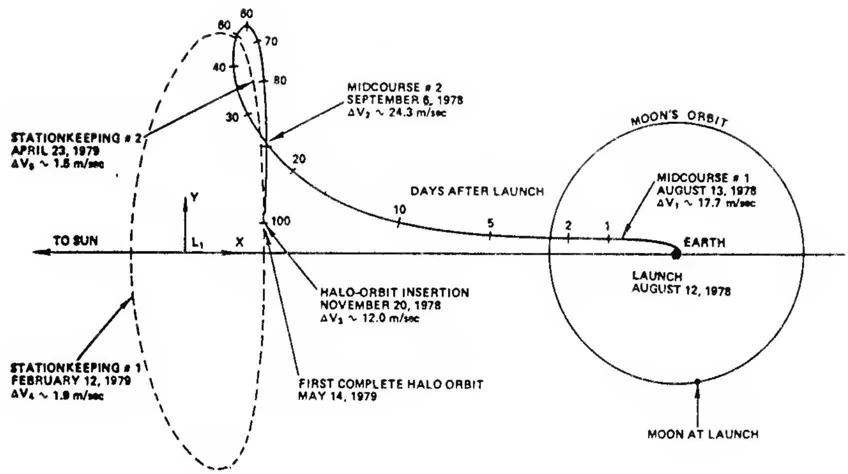


Figure 13 - ISEE-3 Transfer Trajectory to Halo Orbit.

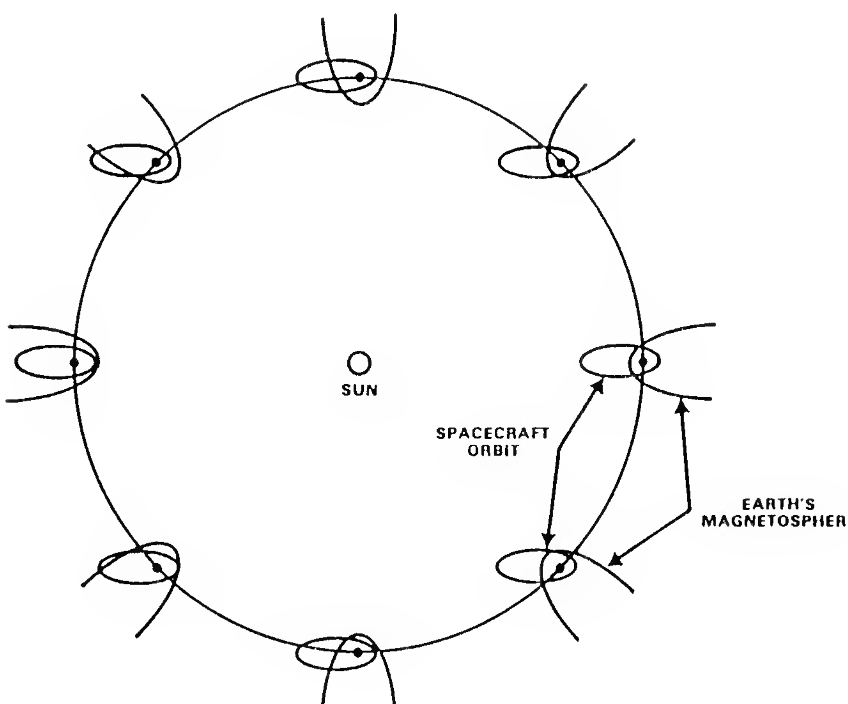


Figure 14 - Uncontrolled Argument of Perigee.

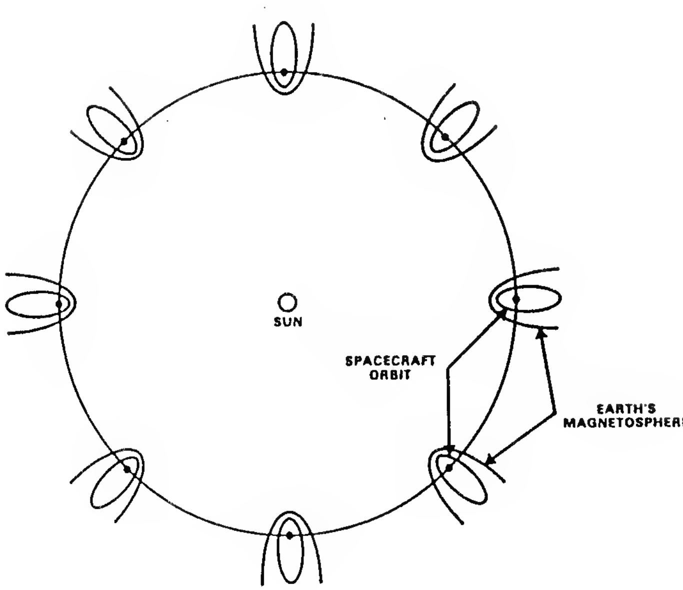


Figure 15 - Controlled Argument of Perigee.

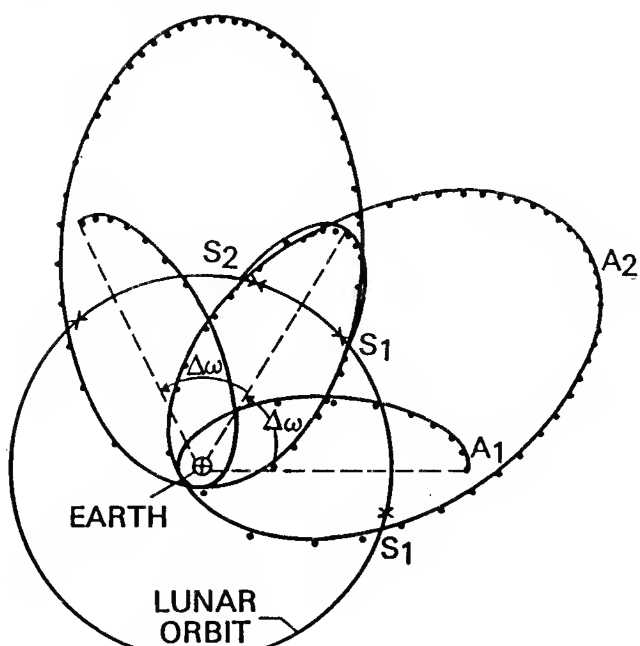


Figure 16 - One-Month Double Lunar-Swingby Orbit, Inertial Frame.

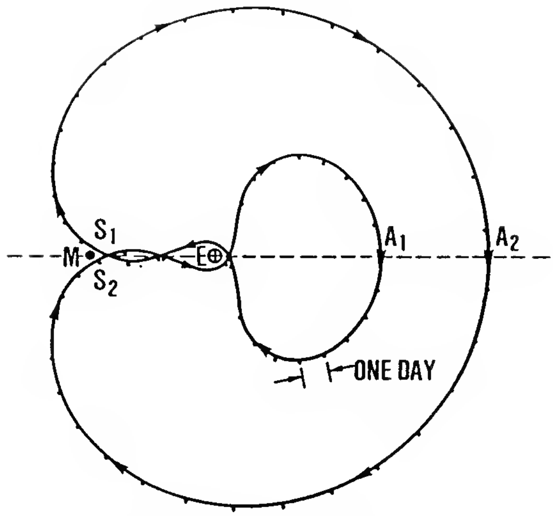


Figure 17 - Double Lunar-Swingby Trajectory in Earth-Moon Reference Frame (One-Month Class).

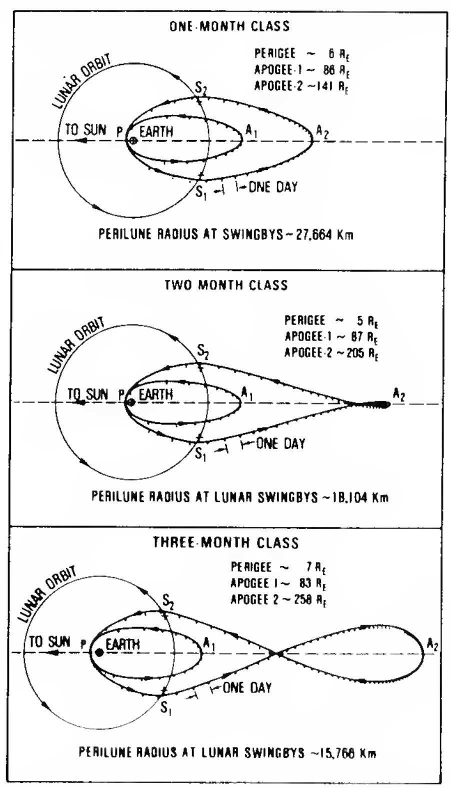


Figure 18 - Sun-Synchronous Periodic Orbits Using Double Lunar-Swingby Technique.

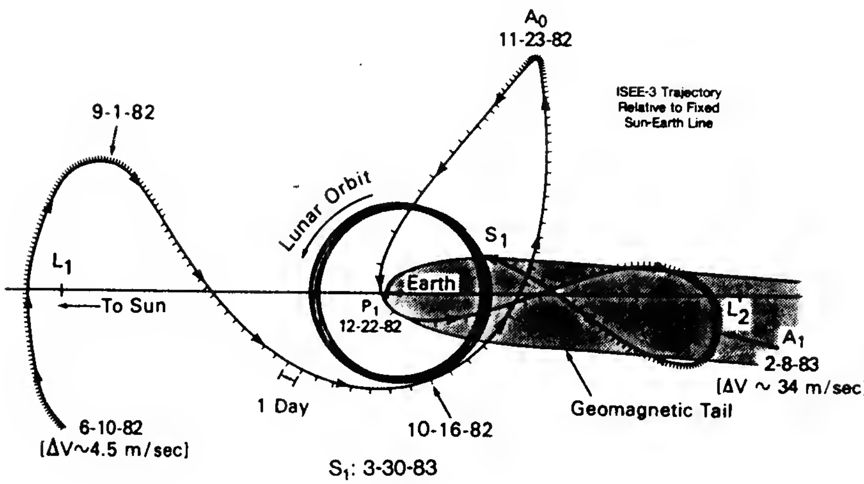


Figure 19 - ISEE-3 Transfer from L1 Halo Orbit to Geomagnetic Tail.

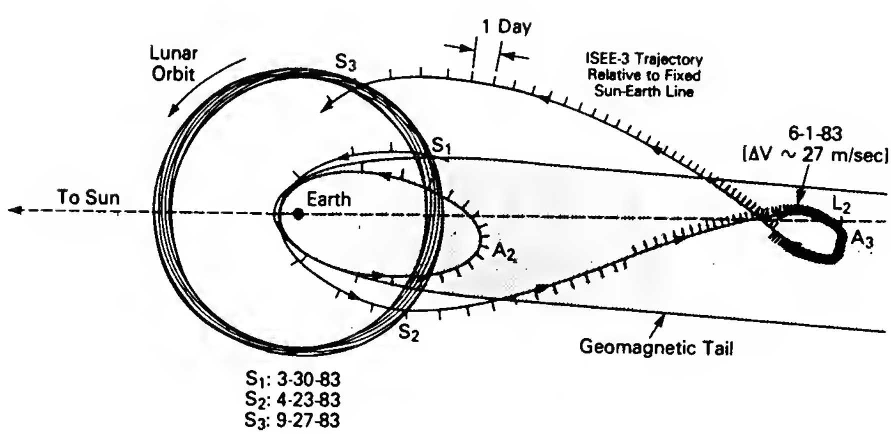


Figure 20 - ISEE-3 Five-Month Geotail Excursion.

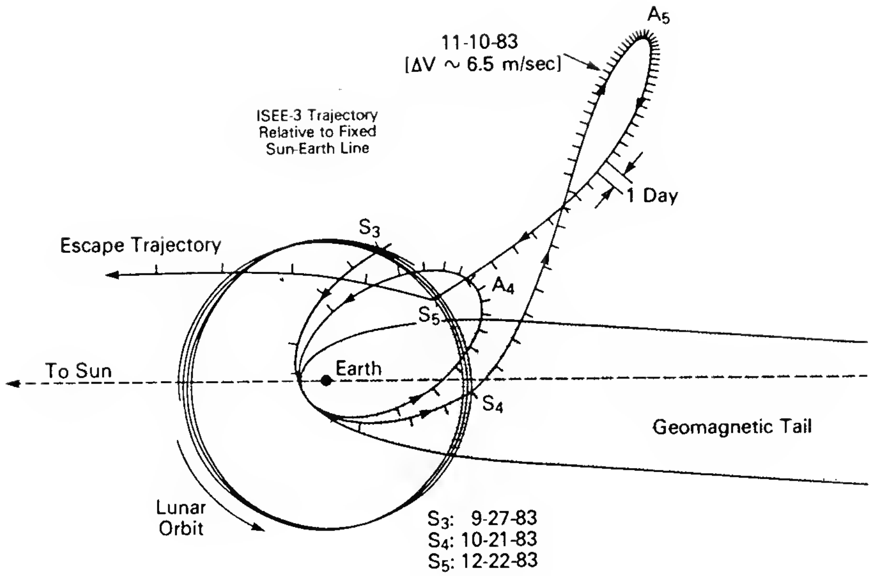
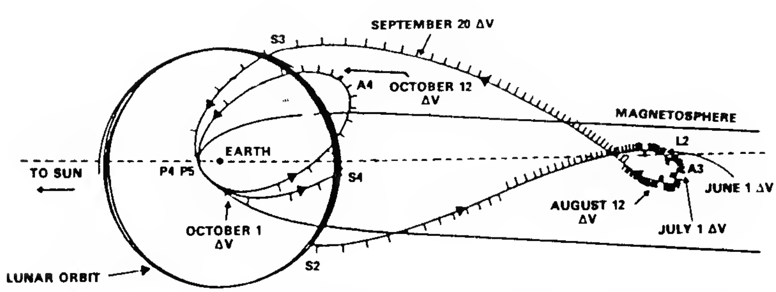


Figure 21 - ISEE-3 Escape Trajectory.



SWINGBY	ORBIT ADJUST ΔV REQUIRED (m/s)	ΔV PERFORMED (m/s)	LUNAR SWINGBY DISTANCE (km)		
			DESIRED	OBSERVED	DIFFERENCE
S3	27.14	27.03	22,464	5,711	30,175
S3	–	2.21	24,514	15,323	9,185
S3	–	0.31	24,508	24,524	16
S3	–	0.03	24,508	24,527	19
S4	–	0.14	19,180	19,211	31
S4	–	0.15	19,180	19,178	2
TOTAL	27.14	29.87			

Figure 22 - ISEE-3 Maneuvers performed between S2 S4.

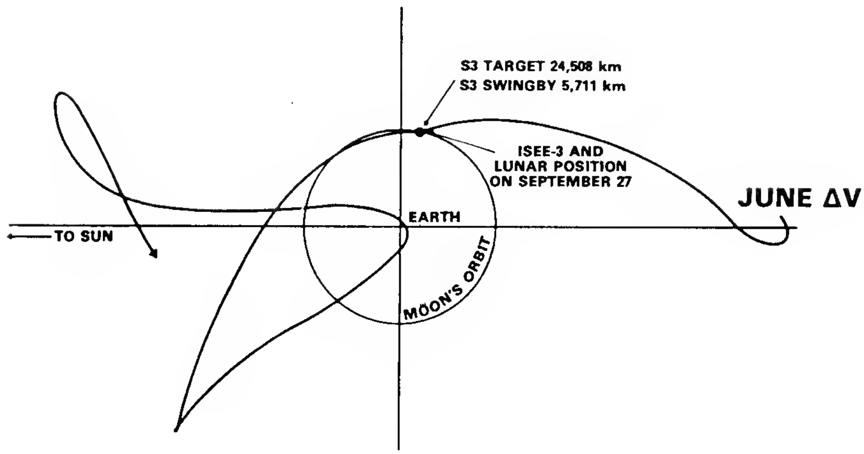


Figure 23 - Trajectory Following June lst Delta-V without Corrections.

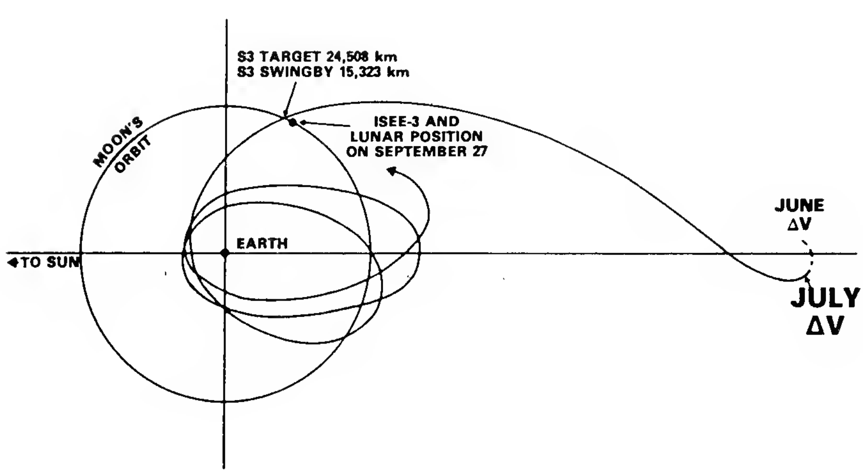


Figure 24 - Trajectory Following July 1st Correction Maneuver.

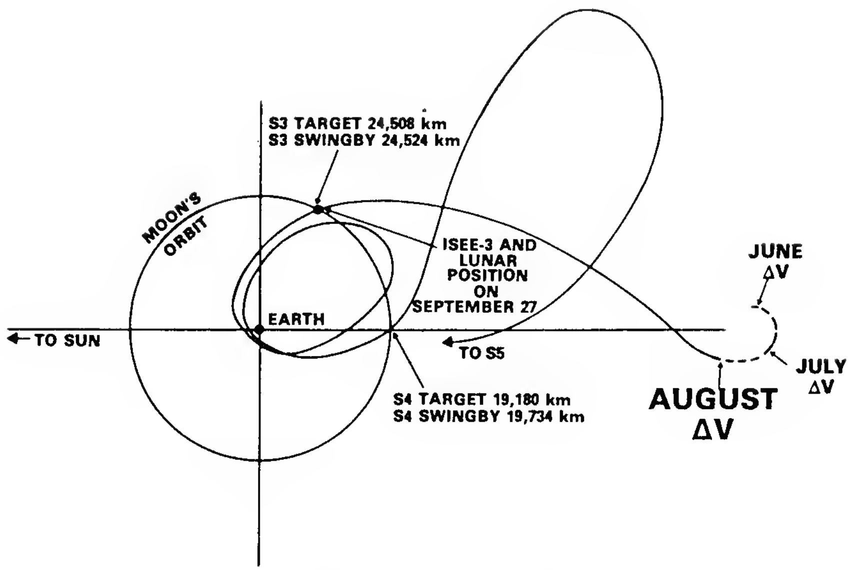


Figure 25 - Trajectory Following August 12th Correction Maneuver.

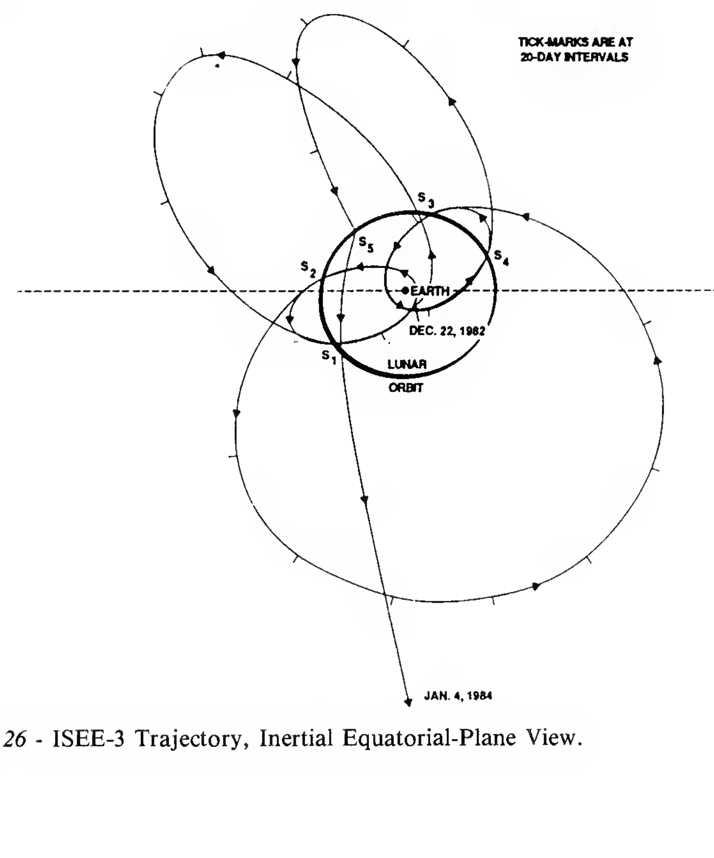


Figure 26 - ISEE-3 Trajectory, Inertial Equatorial-Plane View.

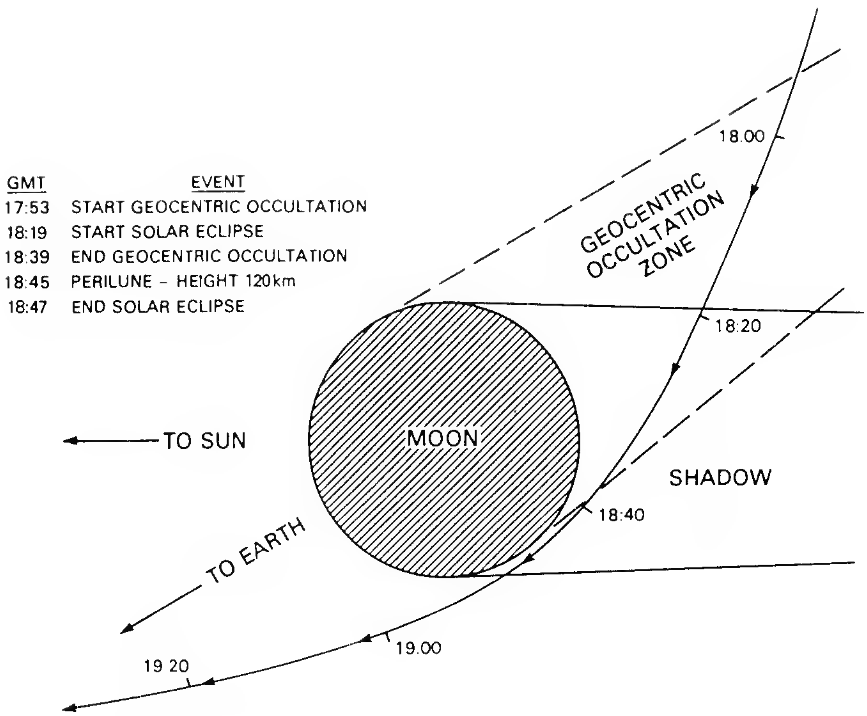


Figure 27 - ISEE-3 Lunar Swingby, December 22, 1983.

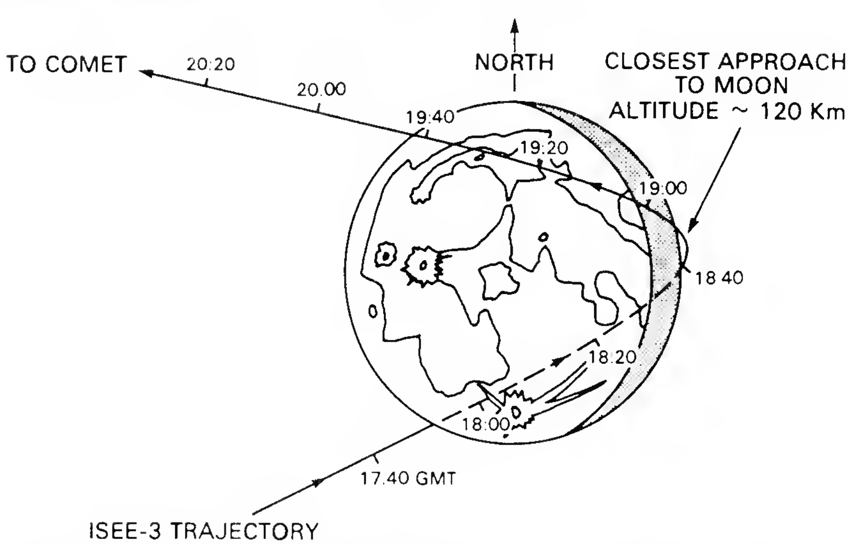


Figure 28 - Trajectory of ISEE-3 around the Moon as seen from the Earth.

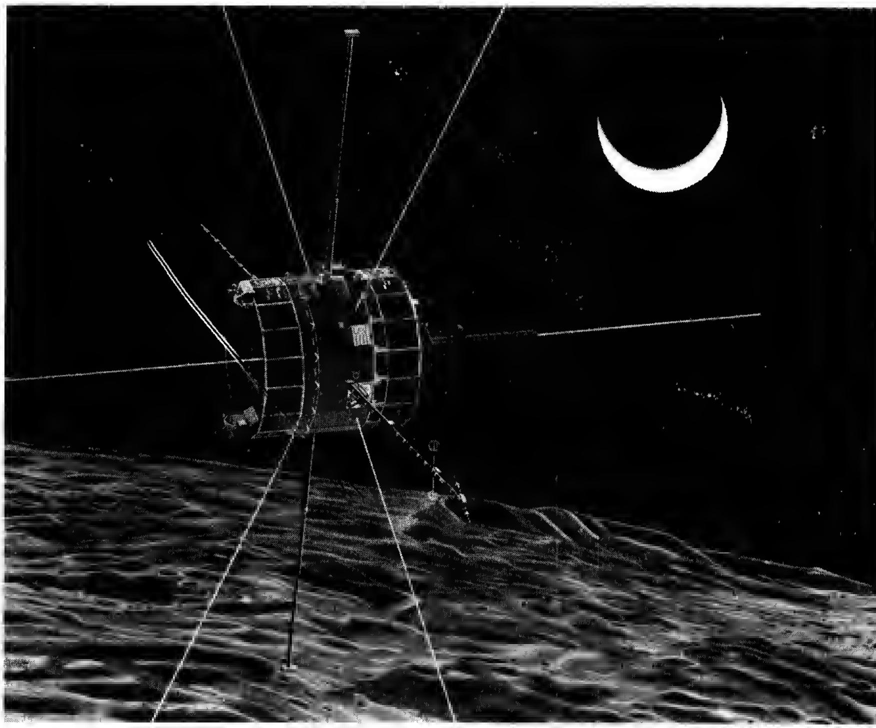
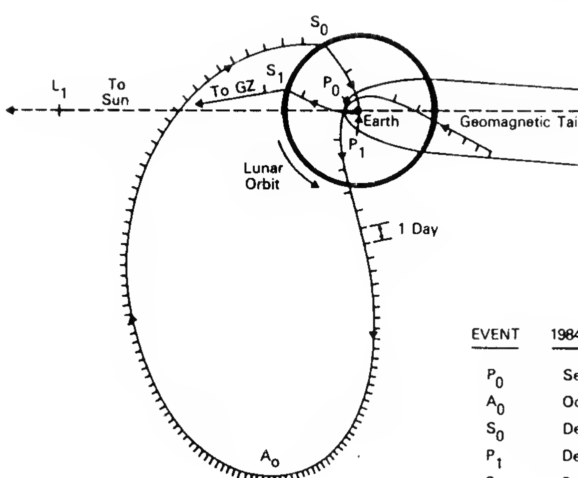


Figure 29 - ISEE-3 Flying 120 kilometers above the Lunar Surface.



EVENT	1984 DATE	DISTANCE	C_3 (km/sec)^2
P_0	Sept. 5	9.4 R_e	
A_0	Oct. 24	307 R_e	
S_0	Dec. 17	2283 km	−0.46
P_1	Dec. 19	1.8 R_e	+1.20
S_1	Dec. 21	1800 km	+4.49

Figure 30 - Hyperbolic Double Lunar-Swingby Escape Trajectory.

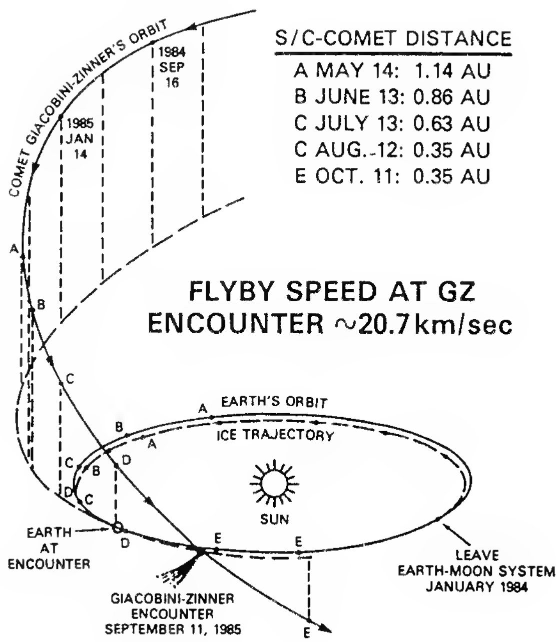


Figure 31 - Trajectories of ICE, the Earth, and Comet GZ in 1985.

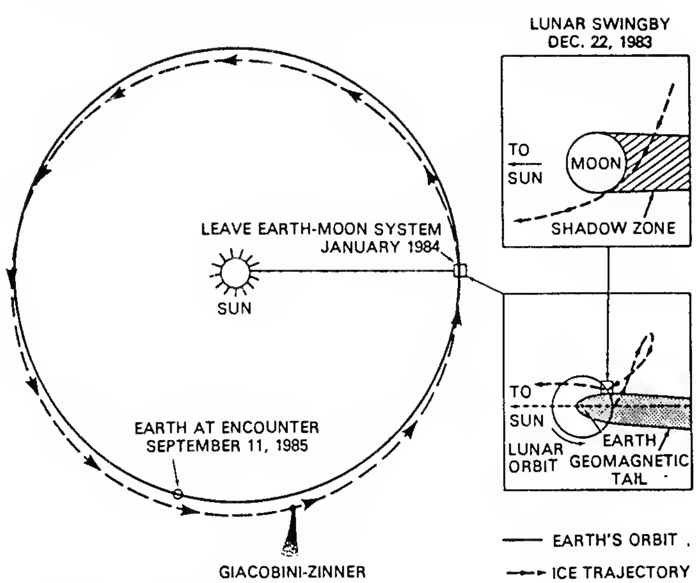


Figure 32 - ICE Comet Mission.

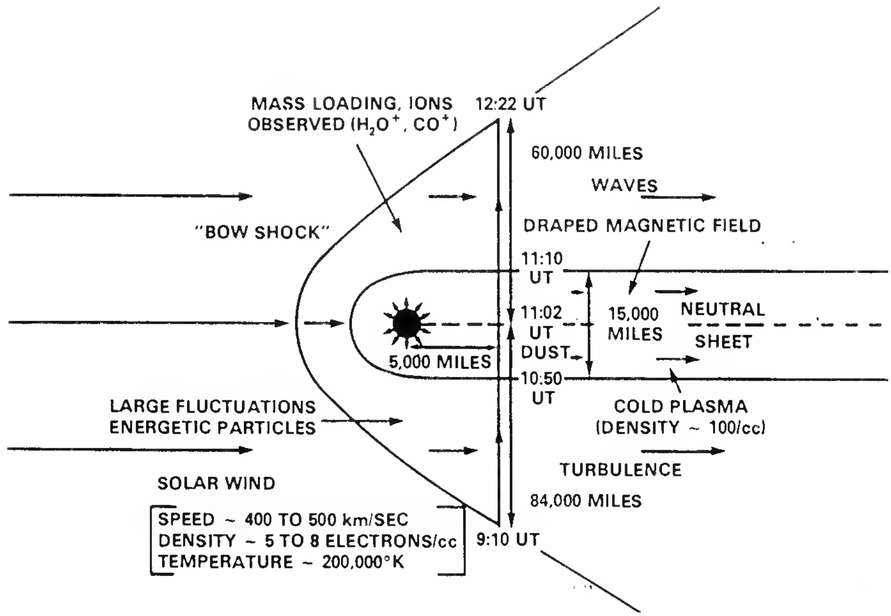


Figure 33 - ICE's Traverse of Comet Giacobini-Zinner's Tail and Coma.

Figure 34 - ICE after Flying through Comet Giacobini-Zinner.

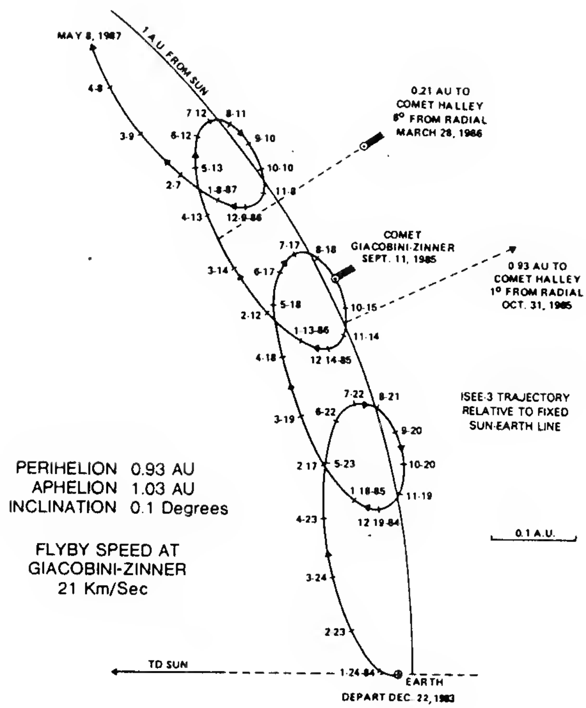


Figure 35 - ICE's Heliocentric Trajectory.

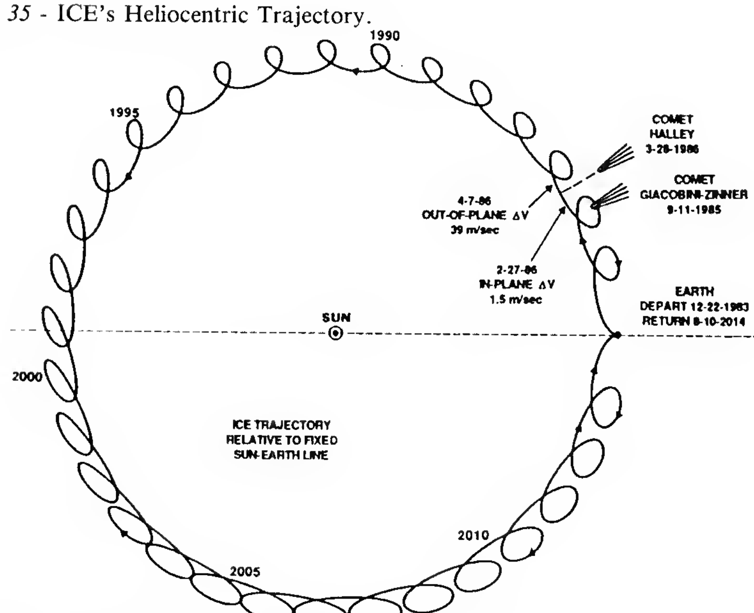


Figure 36 - ICE Earth-Return Trajectory, 1983 to 2014.

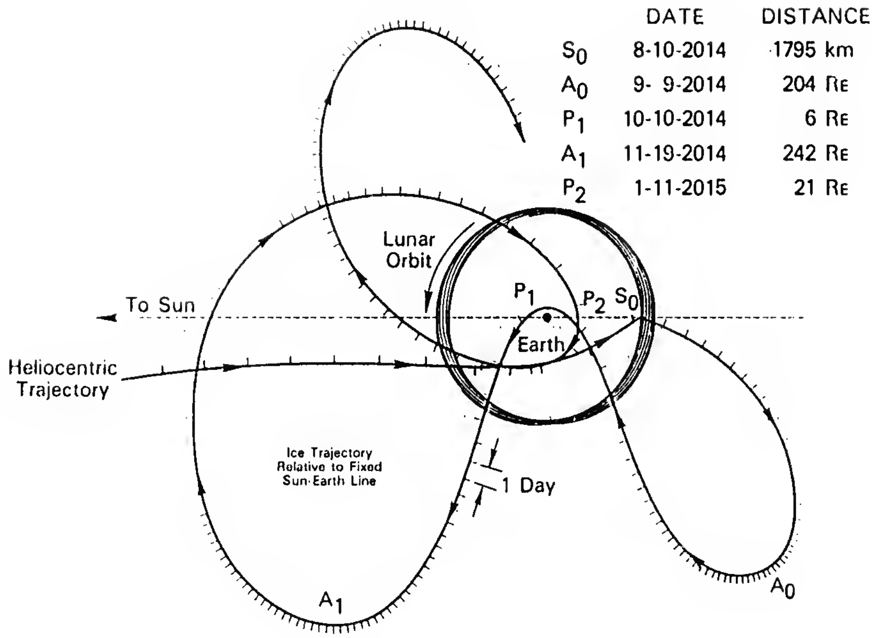


Figure 37 - ICE Capture.

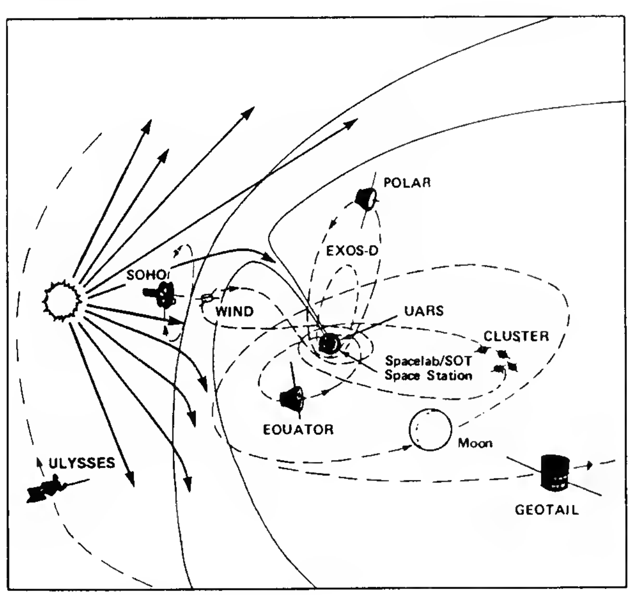


Figure 38 - Orbits of ISTP Spacecraft.

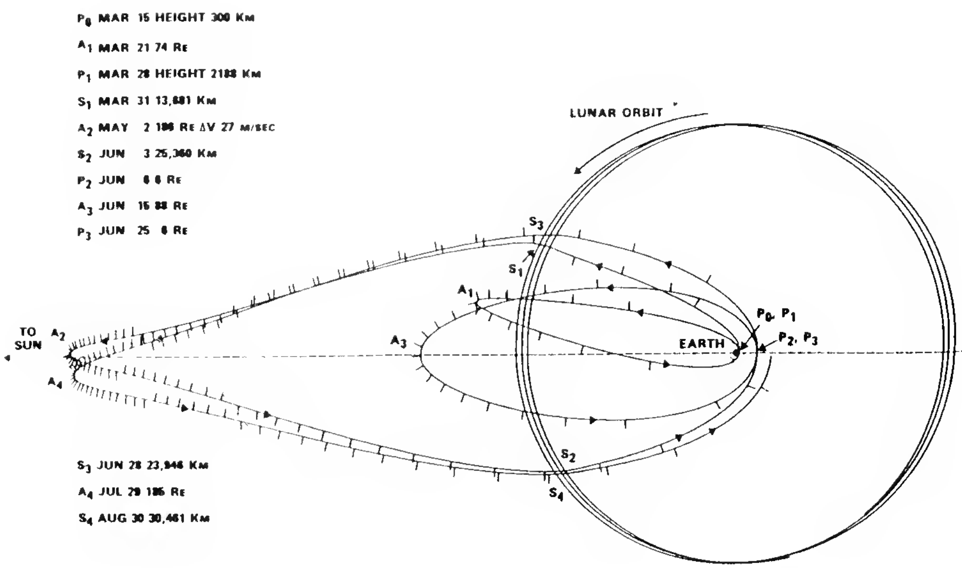


Figure 39 - Wind Spacecraft Sunward Double Lunar-Swingby Trajectory, 1992.

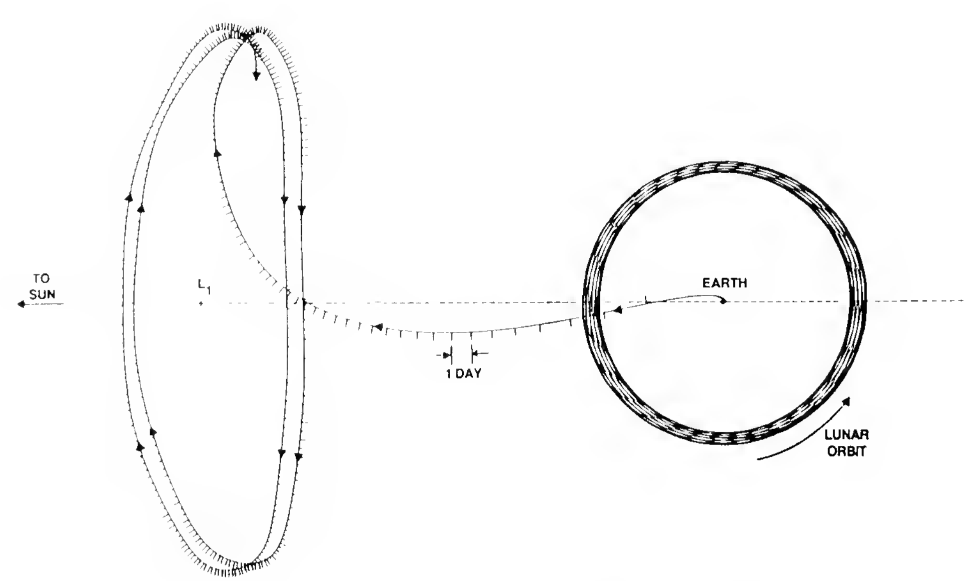


Figure 40 - Direct Transfer to L1 with no Intermediate Delta-V's. 452 Days of the Trajectory beyond the Lunar Orbit are shown.

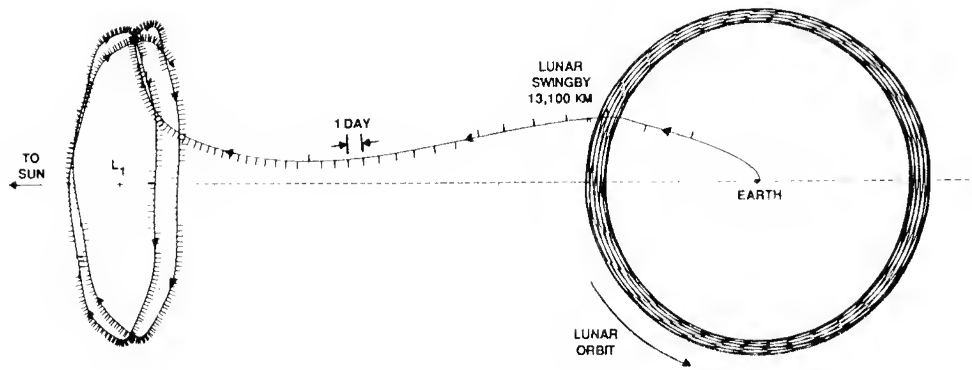


Figure 41 - Wind Transfer to L1 with Lunar Swingby. 443 Days of the Trajectory beyond the Lunar Orbit are shown.

I futuri programmi di Scienza spaziale dell'Agenzia Spaziale Europea

Vittorio MANNO*

I programmi futuri si inquadrano nel Piano a lungo termine denominato: Scienza Spaziale - Orizzonte 2000.

Basato su quattro pietre angolari corrispondenti a quattro progetti maggiori, già identificati, esso comprende inoltre un certo numero di progetti di media grandezza, da scegliersi in competizioni al termine di opportune fasi di studio e l'utilizzazione della piattaforma riutilizzabile Eureca o simili piattaforme scientifiche.

Descriverò nel seguito i progetti già approvati od in fase di studio delle due grandi suddivisioni della Scienza Spaziale: l'esplorazione del sistema solare e l'astrofisica.

1. Esplorazione del Sistema Solare

A sviluppo ultimato ma in attesa del lancio è il progetto *Ulysses*. Tale progetto in cooperazione con la NASA è il reliquato dell'ISPM, nel quale oltre lancio ed operazioni, NASA doveva provvedere una seconda sonda per effettuare misure in correlazione con la sonda dell'ESA. Tale sonda è stata successivamente cancellata: Ulysses sorvolerà i poli del sole, dopo aver percorso una traiettoria intorno al pianeta Giove, il quale lo proietterà con effetto di fionda al di sopra del piano dell'eclittica. Ulysses sarà quindi la prima sonda a raggiungere la terza dimensione e a raggiungere latitudini solari superiori a 70°. Esperimenti europei ed americani misureranno le condizioni fisico-chimiche in questa ancora inesplorata regione dello spazio. Attualmente il lancio di Ulysses è previsto in ottobre 1990, con 7 anni di ritardo rispetto ai piani originari.

Nel futuro due pietre angolari di Orizzonte 2000 sono identificate in questa disciplina dello spazio ed una serie di progetti di classe media da scegliersi in competizione.

* European Space Agency, Parigi.

i) *S.T.S.P.*

La prima pietra angolare a essere realizzata è dedicata allo studio delle interazioni Sole-Terra. Chiamata S.T.S.P., da Solar Terrestri al Science Programme, essa si inquadra in un programma internazionale al quale parteciperanno negli anni '90, la NASA, il Giappone e l'Unione Sovietica.

Nell'attuale configurazione STSP è composto di due progetti. Essi sono SOHO (Solar Helioseismology Observatory) e CLUSTER per lo studio in tre dimensioni dei plasmi interplanetari.

La figura 1 mostra le orbite di SOHO e CLUSTER. SOHO verrà avviato verso il punto di Lagrange L 1, intorno al quale orbiterà, osservatorio permanente del Sole.

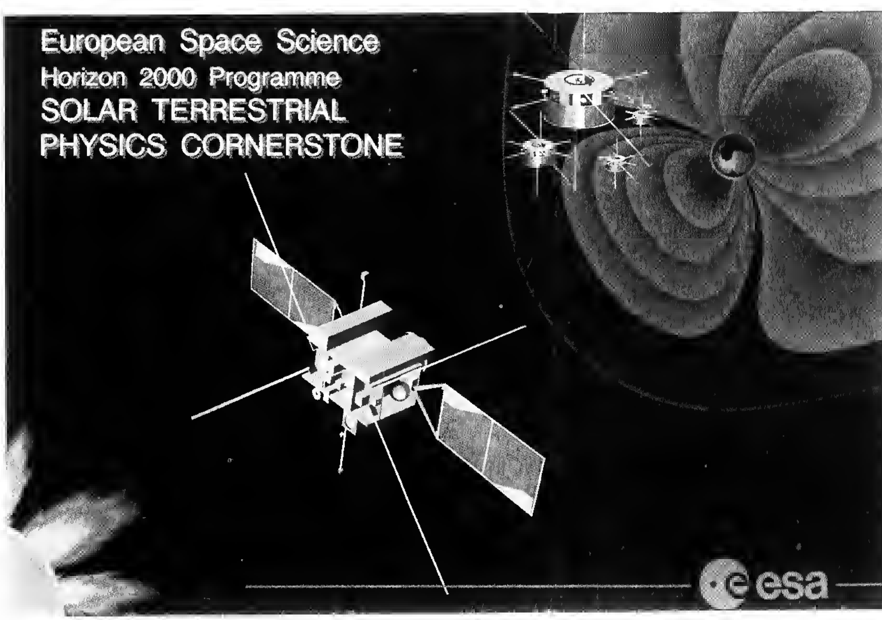


Figura 1 - SOHO e CLUSTER nelle loro orbite spaziali.

A bordo, il carico scientifieo comprenderà:

1) un gruppo di strumenti per la misura dei campi elettromagnetici e delle particelle del vento solare.
2) un gruppo di coronografi per lo studio della corona solare a varie lunghezze d'onda.
3) un gruppo di strumenti per le osservazioni delle oscillazioni solari.

Quest'ultimo gruppo, veramente nuovo nella scienza spaziale, permetterà di trarre informazioni sulla costituzione interna del sole.

Il satellite è stabilizzato su 3 assi con una precisione di circa 1 secondo d'arco.

CLUSTER rappresenta il primo tentativo di effettuare studi dei plasmi extraterrestri in 3 dimensioni. Il progetto consiste quindi di 4 satelliti identici (essenziali per dare la terza dimensione) sui quali saranno imbarcati identici strumenti scientifici.

Le quantità fisiche da misurarsi si estraggono per differenza fra le misure dei 4 satelliti. Di qui nascono delle condizioni molto strette di omogeneità fra i quattro satelliti ed i loro carichi utili.

I quattro satelliti posseggono dei mezzi di propulsione propri, per potere mantenere e variare la distanza relativa a seconda delle caratteristiche dei plasmi nei quali si effettuano le misure. CLUSTER sarà posto in un'orbita ellittica polare che lo porterà attraverso regioni della magnetosfera con plasmi a caratteristiche molto differenti. Le condizioni saranno quindi realizzate per effettuare esperienze di fisica dei plasmi completamente nuove. Inoltre le misure effettuate da CLUSTER nella magnetosfera saranno correlate con le misure di vento solare effettuato dal gruppo di strumenti di plasma a bordo di SOHO.

L'insieme delle misure congiunte e simultanee della corona solare e del vento solare su SOHO e quelle dei plasmi della magnetosfera su CLUSTER, stabiliranno una relazione diretta e quindi un approfondimento della scienza delle relazioni Sole-Terra. La missione S.T.S.P. è presentemente nelle fasi iniziali. Configurazioni alternative sono allo studio. La configurazione di base è basata su una cooperazione stretta con la NASA, la quale provvederebbe al lancio di SOHO nel 1994 ed alle operazioni in orbita. Per contro CLUSTER dovrebbe essere laneiato nel 1995 dal lanciatore europeo Ariane 5.

ii) *Missione solare e cometaria «Post-Giotto»*

Giotto ha compiuto brillantemente la sua missione di esplorazione ravvicinata del nucleo della cometa di Halley. Da questo incontro, Giotto è uscito vacillante ma non distrutto.

Non è dato al momento attuale sapere con precisione la situazione dei sottosistemi e degli strumenti scientifici, in mancanza di una verifica in volo. Ciononostante una possibile nuova missione è stata presentata all'ESA.

Questa consiste nel sorvolo della Cometa Krigg-Skiellerup nel 1992 (una cometa di periodicità di 5.8 anni), riattuando quindi l'esperienza

su Halley, ma su una cometa diversa. L'interesse di questa seconda missione è notevolmente elevato, dandosi la possibilità di effettuare misure su due comete con gli stessi strumenti. Chiaramente la riuscita di tale missione dipende in modo critico dalle condizioni operative della camera a presa d'immagini.

E' allo studio attualmente una fase di verifica tecnica del satellite e degli esperimenti nel 1989 come pre-condizione ad ogni decisione sul futuro di Giotto.

iii) *Le missioni candidate per la selezione del 1988*

Una scelta in competizione di un nuovo progetto scientifico di taglia media o piccola sarà fatta nel 1988. Tre sono le missioni attualmente allo studio nel dominio dello studio del sistema solare che potrebbero presentarsi alla competizione.

Cassini

Questa è una missione congiunta con la NASA, per l'esplorazione di Saturno e del satellite Titano. La Nasa fornirebbe l'orbita di Saturno che provvederebbe anche al trasporto di una sonda dell'ESA per l'esplorazione di Titano, la cui atmosfera presenta una composizione chimica quale avrebbe potuto essere quella della Terra alle origini.

La sonda dell'ESA dopo avere lasciato il veicolo-Madre, dovrebbe entrare ed attraversare l'atmosfera di Titano, durante un periodo di 3 ore, prima di posarsi sulla superficie. Nel contempo il veicolo Madre, inizierà una esplorazione per un periodo di 4 anni, durante i quali studierà il sistema di anelli e tutti i satelliti del pianeta gigante.Le due sonde Madre e Figlia sarebbero lanciate nel 1996, e raggiungerebbero Saturno circa 7 anni più tardi nel 2003. La figura 2 illustra il percorso della sonda dell'ESA attraverso l'atmosfera di Titano.

Vesta

Tale missione è prevista in cooperazione con il CNES francese e con l'Istituto IKI dell'USSR. Si tratta di una missione per il sorvolo e lo studio di un certo numero di asteroidi e di una cometa. Quindi si riprende il filone aperto da Giotto, per lo studio dei corpi primitivi del sistema solare.

Il satellite verrebbe lanciato tramite un lanciatore sovietico. Durante il sorvolo dei penetratori forniti dai sovietici verrebbero lanciati sugli

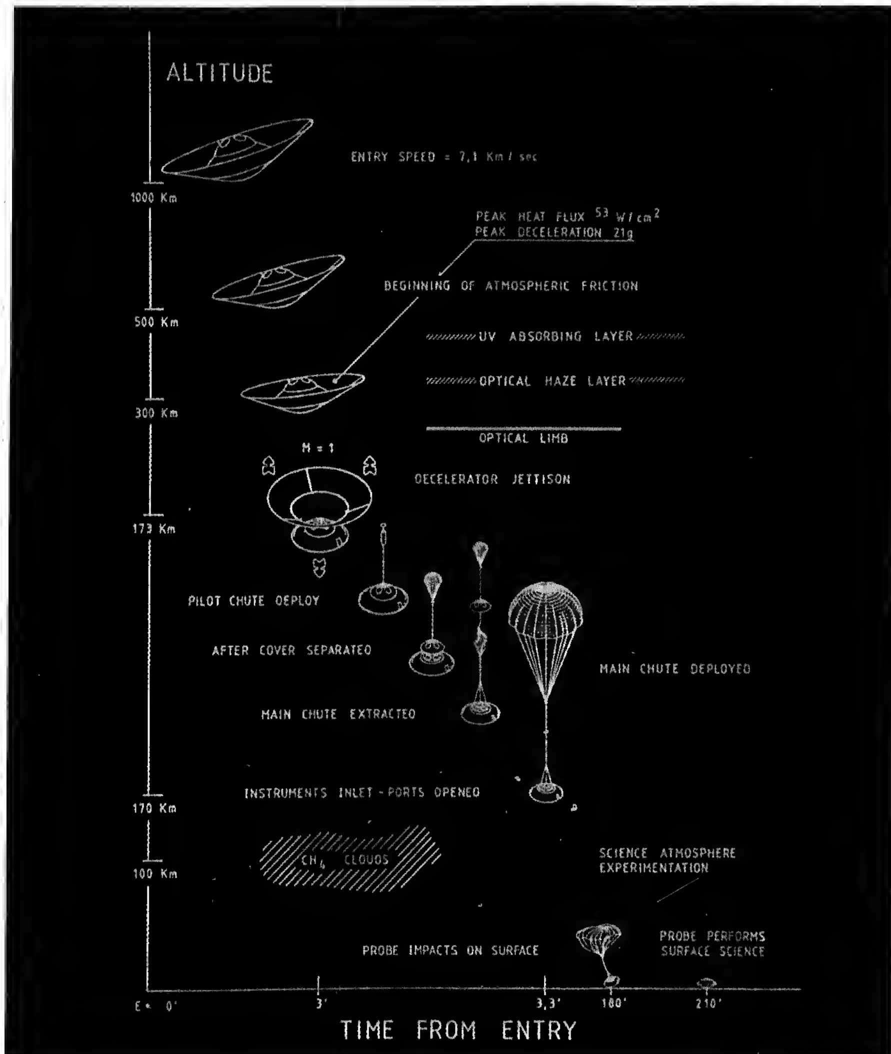


Figura 2 - CASSINI. La sonda ESA attraversa l'atmosfera di Titano.

asteroidi per un'analisi fisico chimica del materiale in situ. A bordo dei sistemi a presa di immagini darebbero una vista complessiva dell'oggetto. Il contributo sovietico verterebbe essenzialmente sul lanciatore e sui penetratori.

Il CNES e l'ESA fornirebbero il satellite stesso e gli strumenti scientifici.

La missione è attualmente in fase di studio di fattibilità e sarà candidata per selezione alla fine del 1988.

La figura 3 mostra le caratteristiche della missione.

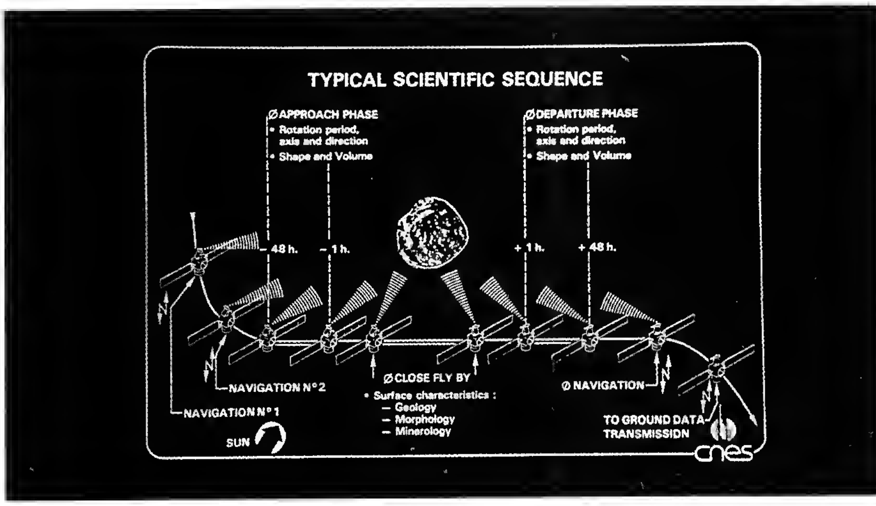


Figura 3 - Caratteristico profilo operazionale di VESTA.

iv) *Un'altra pietra angolare: Comet Nucleus Sample Return*

Questo progetto rappresenta una seconda pietra angolare del programma Orizzonte 2000.

Definita in questo Piano come una missione ai corpi primitivi del sistema solare, è stata in seguito indirizzata all'obbiettivo preciso di rendez vous con una cometa e ritorno a terra di materiale del nucleo cometario per analisi in laboratori terrestri. Si capisce l'importanza di un'analisi chimico-fisica dettagliata del materiale del quale si suppone fosse formata la nebulosa primigenia all'origine del sistema solare. Si capisce anche quanto costosa e tecnicamente avanzata sia tale missione, che di necessità deve essere concepita e condotta in un quadro largamente internazionale.

Al momento attuale studi preparatori sono perseguiti dall'ESA e dalla NASA in parallelo. L'intenzione è di arrivare alla fine del 1987 ad una definizione generale del progetto ed una prima distribuzione di responsabilità fra la NASA e l'ESA.

Gli obbiettivi specifici della missione sono stati definiti in un simposio tenutosi a Canterbury in Inghilterra nel Luglio 1986 (ESA-SP249). L'interesse in Europa per questa missione è intenso in quanto si inserisce

nel diretto filone di ricerca così spettacolarmente aperto da Giotto, ed in quanto promette degli sviluppi tecnologici assai avanzati, particolarmente nel dominio della propulsione elettrica.

La figura 4 dà una rappresentazione artistica della missione.

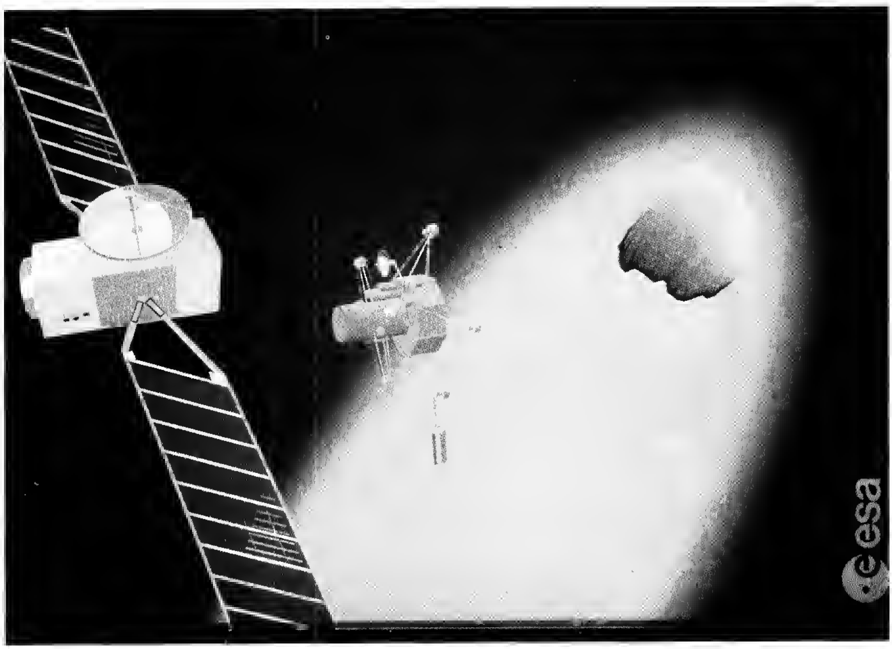


Figura 4 - Comet Nucleus Sample Return.

2. Astronomia ed Astrofisica

In fase di sviluppo attualmente sono:

— HST (Hubble Space Telescope), il grande telescopio di 2.4 m della NASA, al cui sviluppo l'ESA partecipa con i pannelli solari, la camera per oggetti a debole luminosità (FOC) ed un contributo in natura alla fase operazionale. L'ESA otterrà accesso a livello del 15% del tempo di osservazione in favore degli astronomi europei.
Il lancio è previsto nell'Agosto 1989 e la durata di vita operazionale è prevista di oltre 11 anni.

— Hipparcos. Satellite per astrometria, misurerà le coordinate astronomiche di 100.000 stelle distribuite uniformemente sulla sfera celeste, con una precisione di un ordine di grandezza superiore a quanto ottenibile a terra. Il lancio è previsto nel 1989 e la pubblicazione del nuovo catalogo nel 1991.
— ISO (Infrared Space Observatory). Questo primo osservatorio spaziale nell'infrarosso avrà a bordo uno specchio di 60 cm raffreddato e 4 strumenti scientifici per misure fotometriche e spettrali delle sorgenti infrarosse. Il lancio è previsto nel 1993 e la vita operazionale si estenderà fino ad esaurimento del liquido criogenico previsto dopo 18 mesi.

Nel futuro due pietre angolari di Orizzonte 2000 sono dedicate a due progetti maggiori in astronomia, l'uno per lo studio delle sorgenti di raggi X ed il secondo per lo studio nelle lunghezze d'onda mm e sub mm.

Altri progetti di classe media sono allo studio per la selezione in competizione nel 1988.

i) *La pietra angolare:*
Missione di spettroscopia di sorgenti di raggi X

È un progetto di tipo osservatorio che servirà l'intera comunità astronomica alla fine del secolo.

La missione scientifica è stata definita nel Simposio tenutosi a Lyngby in Danimarca nel Giugno 1985 (ESA-SP 239). Essa consiste:

a) spettroscopia in larga banda con massima sensitività (5000 cm^2 a 8 Kev, 10.000 cm^2 a 2)
b) spettroscopia con media risoluzione < 3 Kev (Res: 250)
c) spettroscopia ad alta risoluzione in bande prestabilite (Res > 1000 nelle linee coniche e Fe XXV).

Il carico utile consisterà di 4 telescopi a incidenza radente, composti di 58 coniche ciascuno con lunghezza focale di 8 m.

Nel piano focale saranno posti diversi sensori a stato solido e di tipo bolometrico forniti dai diversi gruppi scientifici.

Il progetto non è ancora iniziato, per quanto una fase industriale preliminare sia in atto, vertente sui particolari problemi tecnologici relativi alla produzione in serie dei telescopi e all'allineamento delle coniche.

Il satellite stabilizzato su 3 assi sarà lanciato in un'orbita ad alta eccentricità (periodo = 48 ore) da un Ariane 4 alla fine del 1998, ed avrà una vita operazionale di circa 10 anni.

La Figura 5, rappresenta schematicamente il progetto.

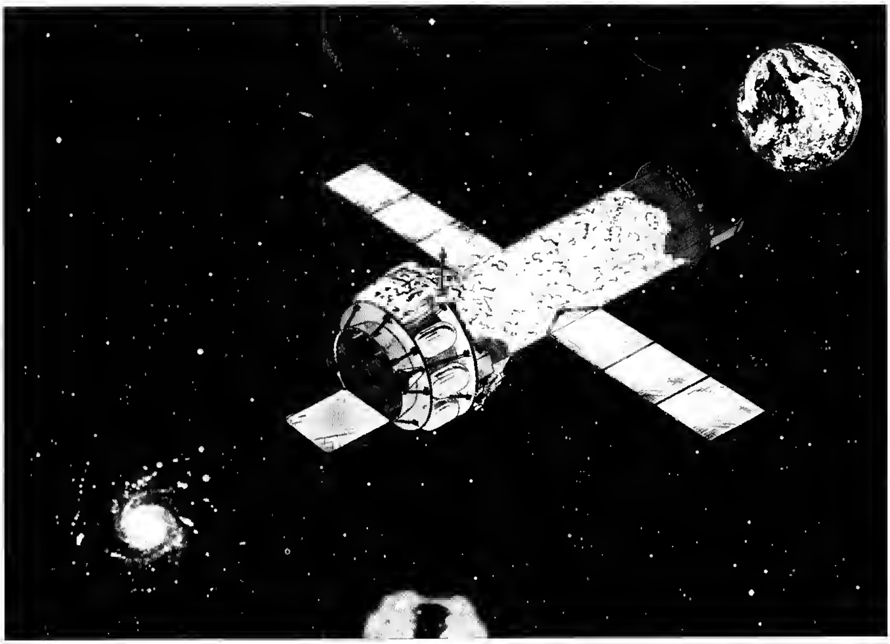


Figura 5 - Missione di spettroscopia di sorgenti di raggi X.

Il progetto sarà gestito come un osservatorio a terra: il tempo d'osservazione sarà in gran parte a disposizione degli astronomi europei, salva una parte riservata agli istituti scientifici che avranno provveduto e finanziato i sensori. Il programma di osservazione sarà basato su proposte di osservazione sollecitate tramite periodiche richieste.

Per quanto approvato, in quanto parte delle pietre angolari, il progetto dettagliato di sviluppo e spesa verrà presentato per approvazione dettagliata al Consiglio Scientifico dell'ESA nel 1991.

ii) *La pietra angolare: spettroscopia di sorgenti nel dominio mm e sub-mm*

Tale progetto consisterà in un telescopio orbitante di 8 m diametro, raffreddato radiativamente a circa 100°K e operante tra 100 μ e 1 mm. È previsto raggiungere una sensibilità di circa 10 MJy.

La definizione scientifica di tale missione è stata fatta al Simposio tenutosi a Segovia nel Giugno 1986 (ESA-SP 260).

Il lancio del Telescopio per mezzo di un lanciatore Ariane non è previsto prima del 2000; tale data è determinata sia da limitazioni finanziarie

sia dalla necessità e opportunità di capitalizzare sui risultati scientifici e sviluppi tecnologici messi in atto per ISO.

Nel campo tecnologico gli studi industriali vertono sullo spiegamento nello spazio della grande autonomia.

iii) *Le missioni candidate per la selezione del 1988*

Lyman

Questo progetto è in qualche modo il continuatore dell'IUE. La missione scientifica principale consiste in spettroscopia ad alta risoluzione e sensibilità nella regione 912 - 1216 A. Tale regione è estremamente ricca di linee spettrali e di informazioni fondamentali. In particolare il progetto determinerà:

1) la misura accurata dell'abbondanza del deuterio nel mezzo interstellare locale e nel mezzo intergalattico a bassi redshift;
2) la mappa dell'OVI interstellare presente nel disco e nel polo della Galassia;
3) la mappa dell'H_2 e HD interstellare nel disco e nell'alone della Galassia.

Il telescopio a incidenza radente ha un diametro di 45 cm, f/10 ed una superficie di 500 cmq. L'orbita preferita è di 120.000 x 2000 km, con periodo di 48 ore.

Il progetto è attualmente in fase di studio di fattibilità e sarà candidato alla selezione di fine 1988.

La figura 6 ne dà una rappresentazione schematica.

Quasat

Il progetto consiste nel lancio di un'antenna di circa 10 m in orbita ellittica, per effettuare misure a larga base (VLBI) di sorgenti radio, in collegamento con le reti VLBI dell'Europa, USSR, USA ed Australia. La risoluzione angolare ottenuta permetterà lo studio dei nuclei galattici e delle quasars su scale coerenti con quelle stimate per i dischi intorno ai buchi neri massivi. Il progetto è attualmente in fase di studio di fattibilità. Collaborazioni internazionali si stanno analizzando con l'Australia, Canada ed anche con l'India e l'Unione Sovietica.

La figura 7 dà una rappresentazione schematica del progetto.

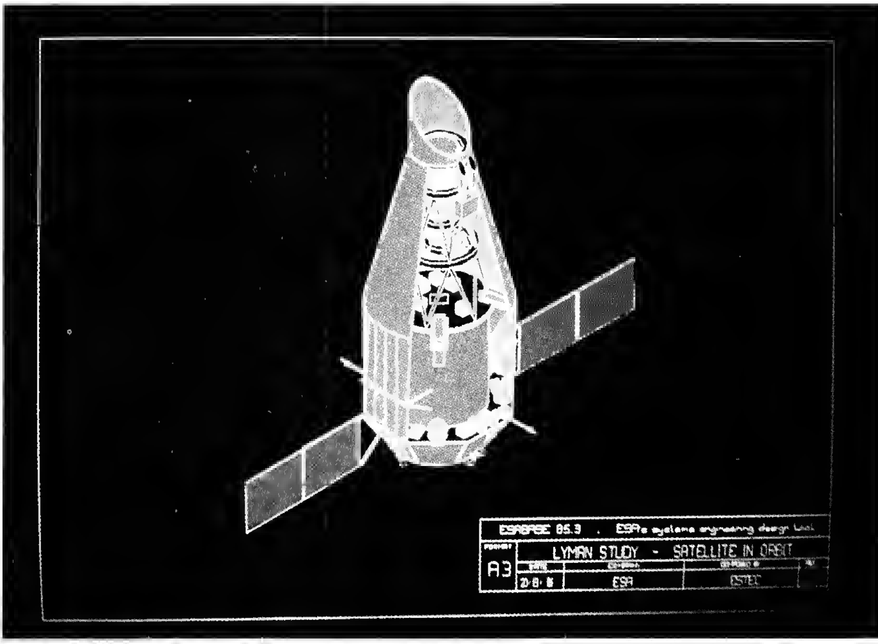


Figura 6 - Lyman: osservatorio nella banda 912-1216 A.

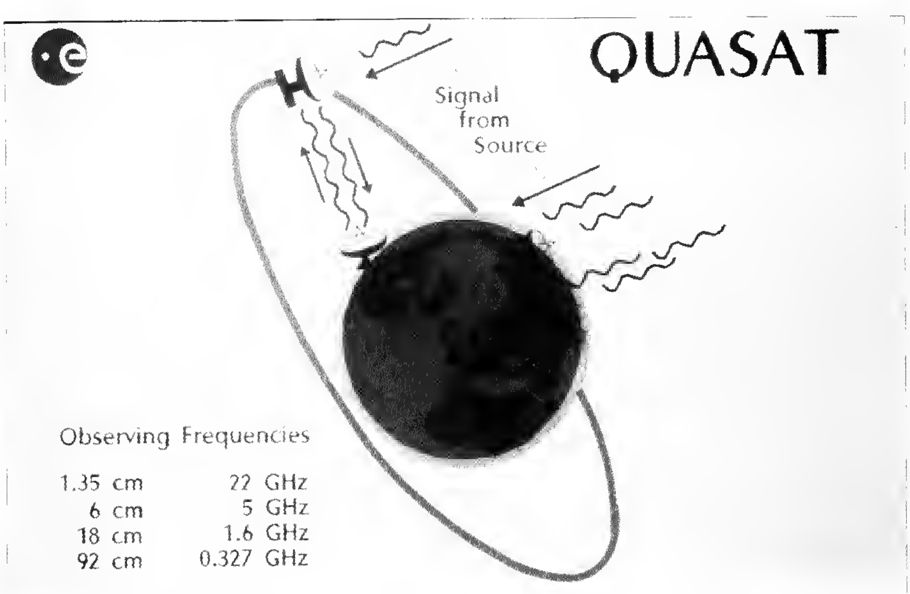


Figura 7 - Quasat: telescopio per misure VLBI.

GRASP (Gamma Ray Astronomy with Spedioscopy and Positioning)

Tale progetto ha come obbiettivi scientifici: misure spettroscopiche con alta risoluzione angolare e con accurata determinazione della posizione di sorgenti di raggi γ. Tale area di ricerca trova molto seguito in Europa in quanto si inserisce nel filone di cos-B.

Il progetto consiste essenzialmente di un telescopio, basato su uno schermo codificato, e su delle matrici di sensori al germanio raffreddato ed al cesio iodato.
Larghezza di banda: 15 Kev a 100 MeV
Risoluzione spettrale: 1000 a 1 Kev (13 Kev - 10 Mev)
Accuratezza di posizione: 1 minuto d'arco
Campo di vista: 5° quadrati
Sensibilità: 10 m Crab (38 in 10^5 secondi)

Il progetto è attualmente in fase di studio di fattibilità.

3. La Stazione Spaziale

Malgrado le incertezze che circondano il concetto e la realizzazione della Stazione Spaziale, il Direttorato prosegue l'analisi di carichi scientifici che trarrebbero particolare vantaggio da questa infrastruttura.

Studi sono in corso tra i quali di particolare interesse è un insieme di strumenti a bordo della stazione stessa di piattaforme coorbitanti o polari e di satelliti per effettuare delle misure congiunte nell'area delle relazioni Sole-Terra (Columbus STP).

Conclusioni

Il programma scientifico futuro basato su Orizzonte 2000, è un programma minimo ma bilanciato, che permette la realizzazione di missioni che coprono l'intero settore di ricerca spaziale europeo, e mantiene nel contempo un delicato equilibrio fra progetti di vasta portata preidentificati e progetti di grandezza media selezionati in maniera competitiva.

Con ciò le necessità di sviluppi di tecnologie avanzate, precondizionali ai progetti maggiori, e di flessibilità del programma, sono salvaguardate. Il programma è iniziato. Per continuarlo e realizzarlo completamente è necessario il sostegno continuo e decisivo dei gruppi scientifici europei e delle autorità politiche e finanziarie degli Stati membri.

Highlights of the Giotto Flyby of Comet Halley

H. Uwe KELLER *

INTRODUCTION

The European Space Agency (ESA) continued to prepare its first interplanetary fligght although NASA of the United States withdrew from the project of a combined cometary mission. The spacecraft Giotto was to encounter comet Halley after perihelion in March 1986. Other space agencies followed. The Soviet Intercosmos redirected two probes (Vega 1 and 2) from Venus and the Japanese sent their first two spacecraft (Suisei and Sagigake) towards comet Halley. The US supported the endeavors by stimulating and coordinating a large ground-based observational program (International Halley Watch). The Giotto spacecraft was the last to encounter the comet during the first minutes of March 14 1986 taking the full risk of destruction by passing only about 600 km from the nucleus. Just before closest approach the attitude of the spinning spacecraft was changed by about half a degree caused by dust impacts. The telemetry link was partly lost for about 20 minutes. The extremely large flyby velocity of 70 km s^{-1} made dust particles appreaciately smaller than 1 g already very dangerous for the spacecraft. A specially designed shield prevented the penetration of dust particles into the interior of the spaceprobe.

The armada of 5 spaeecraft demonstrates our interest in cometary physics and the nature of comets. Comets are quite different from other solar system bodies. They become active and catch our attention only for a short time during their orbits when they come close to the sun. Only a fraction of all comets plunge into the inner solar system from the Oort cloud where comets are stored at 10 to 100 thousand AU from the sun on orbits taking 10^6 y or longer to wander around the sun. The number of comets in this cloud is estimated to be $\approx 10^{12}$. Perturbed by a passing star or a molecular cloud they may become lost to interstellar space rather than fall towards the sun on a very elliptical orbit.

The sudden, unforeseen appearances of comets — sometimes becoming the brightest objects on the sky (next to the moon) for a few weeks

* Max-Planck-Institut für Aeronomie, Lindau Harz.

— have intrigued mankind. At the end of the 17th century it was Sir Edmund Halley who predicted the return of a bright comet with a period of 75 y leading Newton's theory of gravity to a new triumph and showing at the same time that comets are members of the solar system. Halley's comet is in some sense unusual. It is the brightest, most active member of the family of short-period comets (with periods < 200 *y*). Therefore it was best suited for investigation of its activity and interaction with the solar wind by a spaceprobe although its orbit is retrograde. This makes for a high relative velocity between any spaceprobe and the comet.

Our perception of comets has been changed by the results achieved during the comet Halley campaing. The experiments on board the Giotto spacecraft made major contributions. The investigations of the physical (plasma-neutral) processes were an important incentive for the fly-bys. The interaction of the solar wind with the neutral practically bodiless cometary coma (atmosphere) presents a unique situation with analogies in interstellar space that cannot be duplicated in the laboratory. The chemical composition and the micro structure of the nucleus (the comet proper) can be inferred from the coma gas and dust observations. A major new step forward were the first brief observations of the nucleus. The imaging results of the Halley Multicolour Camera[1] will be emphasized in particular.

Solar wind comet interaction

The L_α camera on board the Japanese Suisei spacecraft was the first to observe comet Halley and took images of the extended hydrogen coma of comet Halley. Fluctuations of the central intensity were interpreted as being invoked by the rotation of the nucleus with a period of 2.2 d (Kaneda et al., 1986a; 1986b) similar to a value derived by an analysis of the 1910 ground-based observations (Sekanina and Larson, 1986). The minimum distance of this spacecraft to comet Halley was 150,000 km flying through the bow shock formed by the interaction of the cometary coma with the solar wind. The solar wind passed the comet with a speed of ~ 400 km s^{-1}. Neutral cometary compounds (mostly atoms as final dissociation producs) from the coma were ionized by charge exchange, in this way «mass loading» the solar wind upstream since protons are substituted by heavy ions such as C^+ or O^+.

[1] G. Colombo was member of the team of investigators of the Halley Multicolour Camera. He supported the experiment from its beginning anticipating the achieved breakthrough.

The solar wind was slowed down and transgressed to subsonic speed at the bow shock about 10^6 km upstream of the central coma. The diversion of the solar wind around the «obstacle» coma was well demonstrated by the Japanese observations (Fig. 1). The former cometary atoms

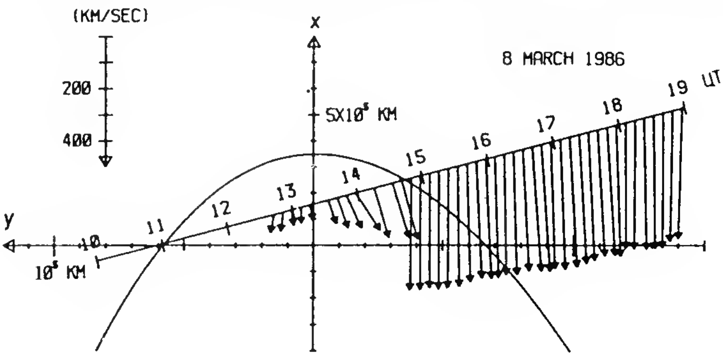


Fig. 1 - Solar wind flow vectors obtained during the encounter of the Japanese spacecraft Suisei with comet Halley on March 6, 1986. The flow vectors and angles are represented in the rest frame of the comet (from Mukai et al. (1986)).

were coupled to thc solar wind with high energie (~ 10^5 eV). These pick up ions made the solar wind flow turbulent and were observed at distances many Gm away from the comet (Gringauz et al., 1986; Johnston et al., 1986).

Further inward the speed decreased more and more reaching the stagnation point around 10^5 km. The decrease of the flow speed of the solar wind protons was also reflected in the decrease of the energy of the pick up ions. The ions picked up within the coma were of even lower energy and formed a separate branch in the energy versus ion mass diagram (see Fig. 2). The two separate populations (solar wind ions and

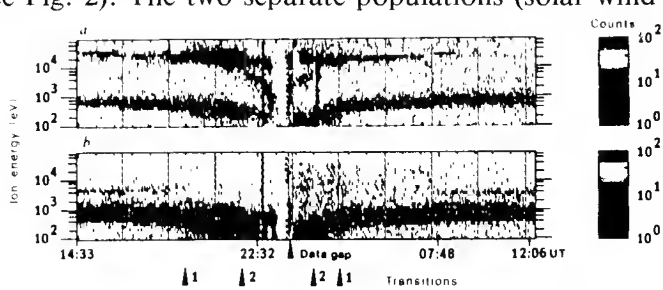


Fig. 2 - Data for two mass groups obtained by one IIS time-of-flight sensor (Johnston et al., 1986). The two mass groups are protons (mass 1 AMU) from solar wind (lower panel) and implanted cometary ions of mass 12-22 AMU (upper panel). Disregard the 'ghost' of the proton signal.

pick up ions) converged. The outflowing cometery ions were slowed down to almost zero speed at about $2 \cdot 10^4$ km producing a secondary maximum in the ion density of the coma (Fig. 3).

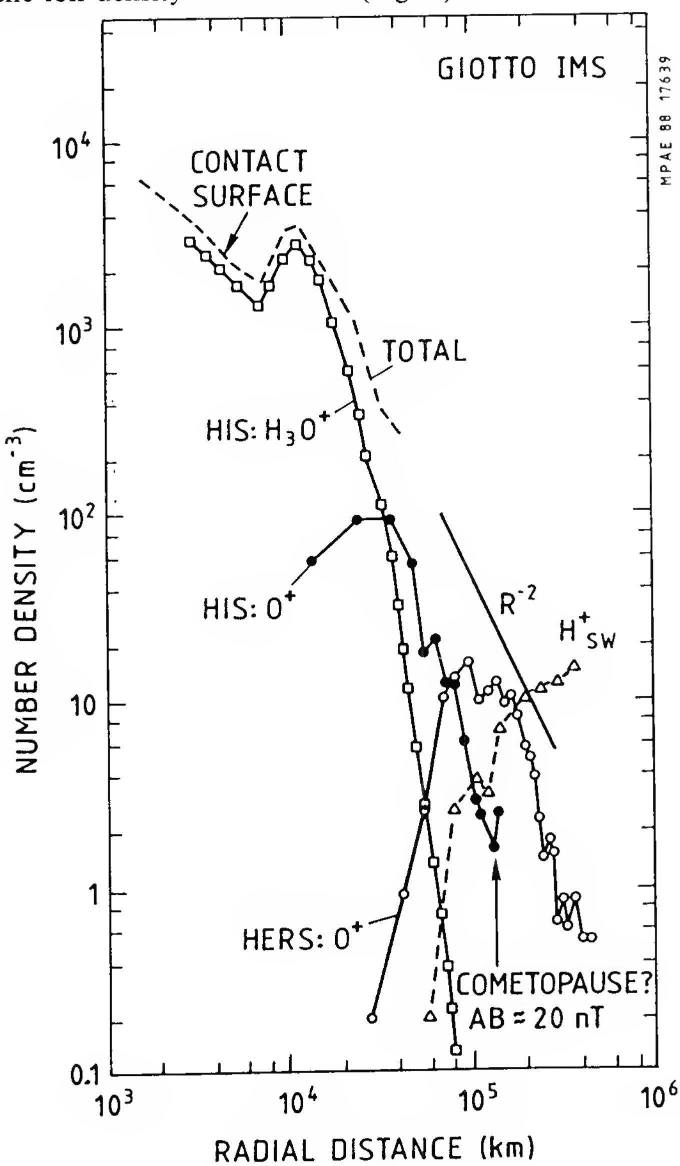


Fig. 3 - A summary diagram of the number density profiles of several water-group ions as observed by the two sensors of the ion mass spectrometer experiment on Giotto (Ip, 1988): solar wind protons and hot O^+ ions (HERS) and cold O^+ and H_3O^+ ions (HIS).

Further in, the Giotto spacecraft detected a magnetic field free cavity when it crossed the contact surface (Fig. 4) at a distance of 4500 km.

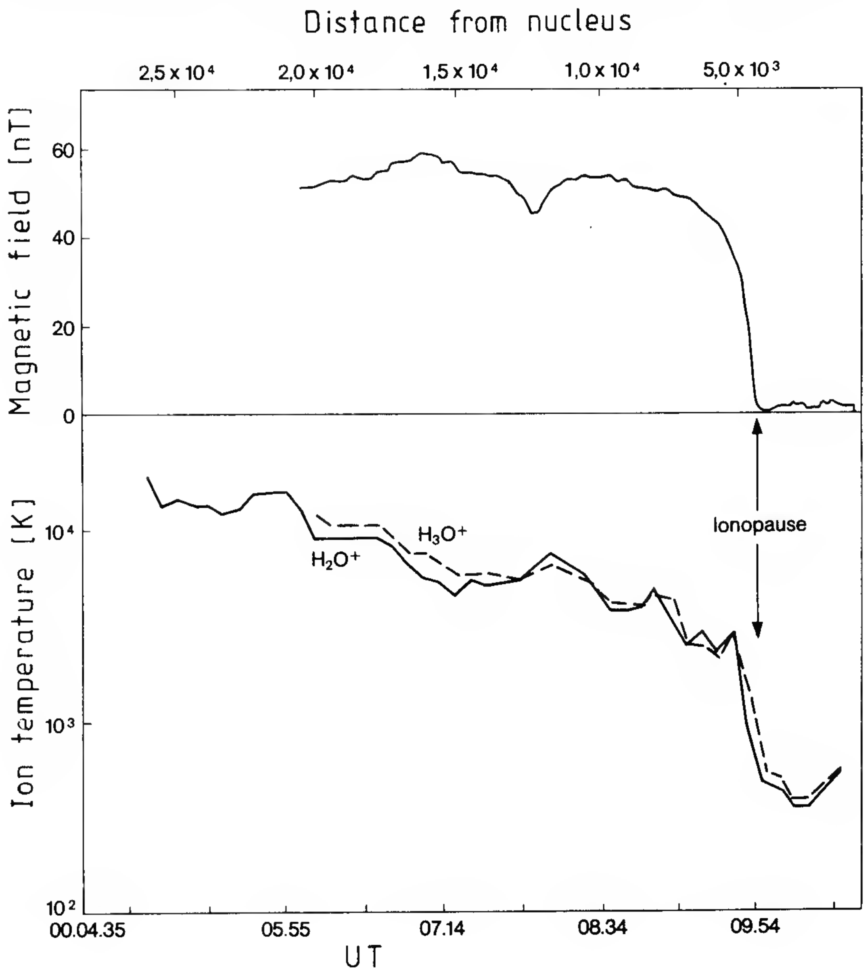


Fig. 4 - The piled up magnetic field drops to zero within less than 30 km at the contact surface or ionopause (upper graph). The ion temperature shows also a strong discontinuity (lower graph) (from Ip (1988)).

The magnetic field dropped from the piled up value of ~ 60 nT down to zero within a fraction of a second. Such a sharp discontinuity was not expected. The size of this cavity came also a surprise. The large stand off distance could be explained by taking the pressure of the neutrals into account to balance the magnetic field force. The ions were still collisionally coupled to the neutrals at distances of several thousand kilo-

meters. The importance of this coupling had not been appreciated.

The ion temperature dropped strongly inside the cavity. Here were only cometary ions, the solar wind could not penetrate. The ion mass spectrometer received high resolution data as close in as 1500 km from the nucleus.

Several layers and discontinuities in the ion distribution have been detected and discussed during the analysis of the data. They were less marked and also less well understood in their significance if compared to the contact surface and the bow shock (see Galeev (1987) for a review).

The cometary coma

The composition of the coma revealed little surprise. The dominating parent molecule (80%) was water H_2O, CO followed with 5-15% and then CO_2 with 3.5%. All other species were in the range of percents and below. Often only upper limits were determined (Table 1). A

Table I

Species	Gas Production Rate Relative to H_20	Remarks
CO	0.05...0.15	Also rocket UV observations
CO_2	$\leqslant 0.035$	
CH_4	$\leqslant 0.07$	
	≈ 0.02	Giotto Ion Mass Spectrometer
NH_3	$\leqslant 0.1$	Gas spectra
	0.01...0.02	Ion spectra
N_2	$\leqslant 0.02$	Gas spectra
	< 0.02	Ion spectra

Table 1 - Abundances of parent molecules in the coma of comet Halley (Krankowsky and Eberhardt, 1988). The data were obtained by the Neutral Mass Spectrometer (NMS) on board Giotto at a heliocentric distance of 0.89 AU.

major part of the observed CO probably came from dust grains that formed an extended source around the nucleus (Krankowsky et al., 1986; Eberhardt et al., 1987). Very volatile compounds such as N_2, CH_4, and also CO (if only the fraction in the ice is considered) were underrepresented. There are several possible explanations. The temperature during formation of the nucleus (condensation of the volatiles and coagulation of the grains) was too high to freeze these highly volatile species. They did not exist in the solar nebula. They had been depleted in the external layers of the nucleus. The temperature of an ice covered surface reaches about 200 K and the gradient into the nucleus causes heat transport into the nucleus. Therefore the outer layers will be much warmer than the temperature in the Oort cloud ($\approx$ 10 K). If the nucleus structure is porous enough diffusion, and with it depletion, of highly volatile compounds seems plausible. This case also sets a big caveat. The composition observed in the coma at a certain point on the orbit and in time does not necessarily reflect the average or original composition of the nucleus.

Heavy molecular ions were detected ranging beyond 100 AMU (Korth et al., 1986). Relative density maxima were found at alternate mass differences of 14 and 16 AMU (Fig. 5). Interpreting these differences as

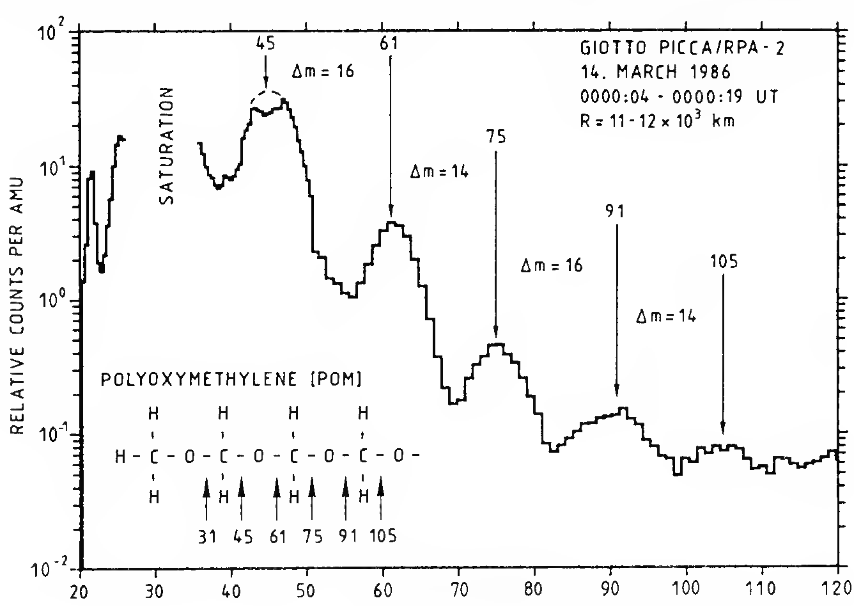


Fig. 5 - Ion mass spectrum (Huebner et al., 1987) at a distance between 11 and $12 \cdot 10^6$ m. The repetitive peaks could be interpreted as dissociation production of the chain molecule polyoxymethylene.

CH_2 and O splitting alternatively from a chain molecule leads to polyoxymethelene as large parent molecule (Huebner, 1987). Other interpretations are possible, even invoking ring structures (Korth et al., 1988). Compounds in question are not easy to sublime and need higher temperatures than that of subliming water ice. These organic tar-like materials probably sublime from the non-volatile dust component either from the grains or to a lesser degree from the hot (otherwise inert) surface.

Dust

The result of the dust grain observations returned very exiting data both from the measurements of the grain size distribution as well as from the determinations of the composition. For the first time it was possible to measure the grain size distribution within the coma directly rather than to infer it from the scattering properties. And sure enough new particles too tiny to reflect the light in the visible were detected. Rather than peaking around a radius of about 0.1 μm the number of small particles kept increasing beyond the limit of the measurements at the very low value of about 10^{-16}g (McDonnell et al., 1987) (Fig. 6). There was a fluid tran-

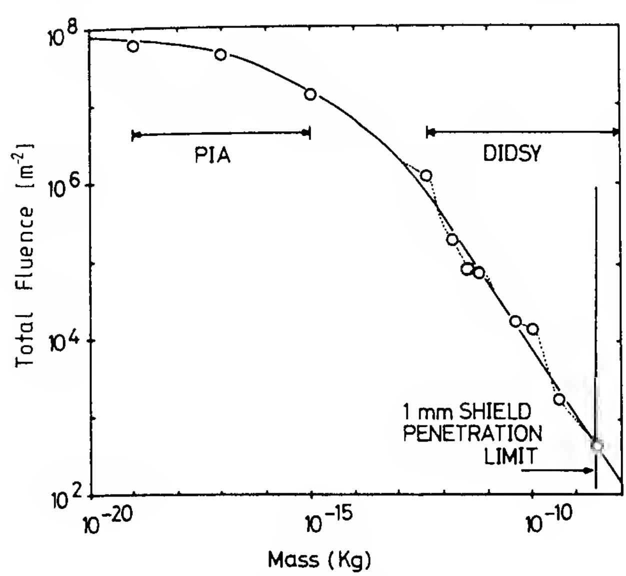


Fig. 6 - Dust particle fluences derived from - 5 min to + 5 min around Giotto closest approach. Data were taken by the instruments PIA and DIDSY (from McDonnell et al. (1987)).

sition from large centimeter size grains to molecular clusters. The main mass of the dust was probably contained in the large particles that could not be observed because of the poor statistics and the masking by impacts of the numerous small grains. In fact the largest impact (about 100 mg) was recorded by the Halley Multicolour Camera registering an attitude change of the Giotto spacecraft.

The spatial dust distribution was not isotropic but concentrated in jets, therefore it was difficult to extrapolate from measurements along the spacecraft trajectories to the total dust production. The dust to gas ratio by mass was estimated to lie between 0.2 and more than 6 (Crifo, 1987). This value is higher than assumed in the past. A 1:1 ratio seems plausible supported also by arguments based on the overall element abundances in comets (Jessberger et al., 1988).

Some of the most exciting results came from the direct measurements of the composition of dust grains in situ. The sophisticated Particulate Impact Analyzer (PIA called PUMA on board the Vega spacecraft- analyzed thousands of particles mainly of small sizes in the range below 10^{-12} g (Kissel et al., 1986a; 1986b). The composition of the grains was found to be highly variable (Clark et al., 1987) ranging from mineralic type grains to mass spectra only showing the volatile elements H, C, O and N (Fig. 7). These CHON particles, contributing one third of the dust mass, constituted one of the biggest surprises. Cometary dust grains contained a large component of organicc material that did not vaporize at the sublimation temperature of water around 200 K. However, when the dust grains left the matrix of the nucleus the small particles were heated to temperatures up to 500 K and the semi-volatile molecules transgress into the gas phase. The high mass ions of organic material found in the inner coma and the appearance of CN and C_2 jets in ground based observations of the coma (A'Hearn et al., 1986) as well as the inferred extended source for CO molecules (Eberhardt et al., 1986) support the idea that the dust grains contribute substantially to the gas component of the cometary coma. The dust grains did not only show a continuous transition from large centimeter size grains to dimensions of molecular clusters but also from non-volatile to volatile material.

Nucleus

The nucleus of comet Halley was observed by cameras on board the Vega (Sagdeev et al., 1986a) and Giotto (Keller et al., 1986) spacecraft. Only the Giotto Halley Multicolour Camera (HMC) revealed details of

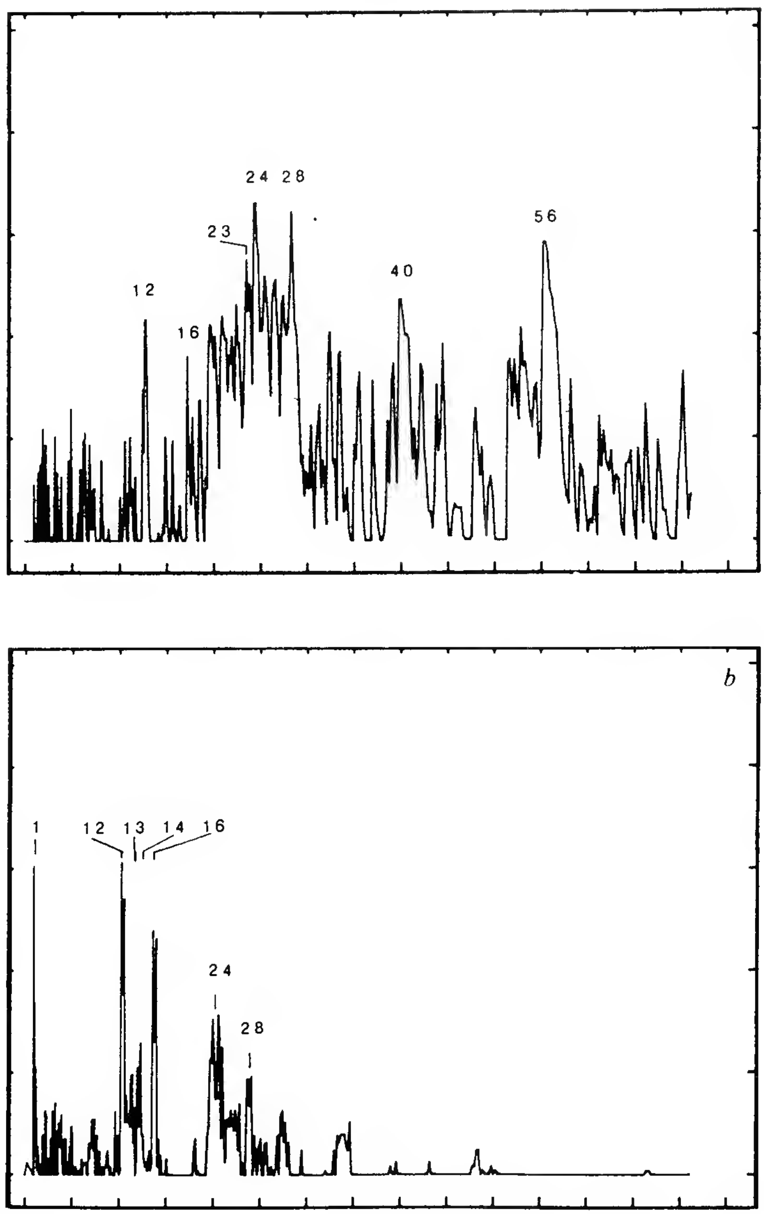


Fig. 7 - Two very different dust composition spectra achieved during the Vega 1 flyby. (a) spectrum closely related to a particle of type CI carbonaceous chondrite; (b) spectrum dominated by light elements, a so called «CHON» particle (from Kissel et al. (1986a)).

the surface morphology and topography. The Giotto spacecraft passed the nucleus at a distance of 596 km in the first few minutes of 14 March 1986 (Curdt et al., 1988). It lost contact with the earth shortly before closest approach.

Therefore, the images taken by HMC during the fly-by covered a distance range from 770,000 down to 1600 km (Fig. 8). The best resolu-

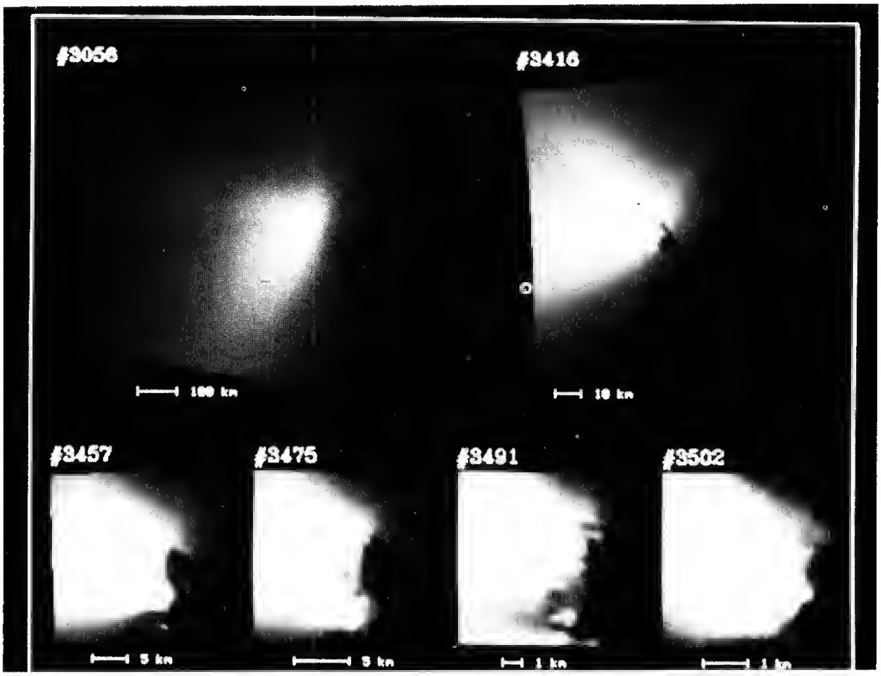


Fig. 8 - Six examples of images of comet Halley in original frame sizes taken by the Halley Multicolour Camera on board the Giotto spacecraft. Image # 3056 (distance to nucleus 124,000 km) was taken 1814 s and image # 3502 31 s (2200 km) before closest approach.

tion was about 40 m per picture element (22 μrad/pixel). The phase angle during approach was 107° and changed only by a few degrees as long as HMC operated. The camera and its operation were described by Keller et al. (1987a).

Not surprising comet Halley was found to have **one** solid nucleus. It was rather irregularly shaped and appreciably larger than expected. Its dimensions are about 16 km × 8 km × 8 km, highly elongated. It had a uniformly dark surface with a geometric albedo of less than 4%. The dark surface showed only localized activity of dust production (jets). Its temperature (> 300 K) was markedly higher than the equilibrium

temperature for sublimating water ice (Emerich et al., 1987). The total surface of the nucleus was about 400 km^2 and its volume about 550 km^3 with an uncertainty of 30% (Keller et al., 1987b). The mass of the nucleus can be estimated from the effect of the 'non-gravitional forces' caused by the non uniform sublimation of the ice near the sun (see Whipple (1986) for a discussion). The derived density of the body is rather low, ≤ 0.5.

Shape of nucleus

The whole outline of the nucleus was visible (Fig. 9); the major part

Fig. 9 - A composite of 6 HMC images ranging in resolution from 320 m to 60 m per pixel. Some of the contributing single images are displayed in Fig. 8.

(75%) as dark silhouette against the scattered light of the dust in the background. The highly elongated shape (2:1) of the nucleus of comet Halley may not be uncommon for comets (e.g. comet Iras-Araki-Alkock, Sekanina (1988)). The sublimation rate (see below) is too small to yield

such an irregular body starting from an approximate sphere. Therefore the nucleus of comet Halley was formed as a highly elongated body. This and the observed coarse roughness support the assumption that the nucleus was formed from rather large subnuclei (dimensions 1/2 to 1 km).

A (relatively recent) break-up of a much larger body could also yield this strongly elongated shape. The almost straight limb on the dark side is suggestive for such an interpretation.

Surface properties

The reflectivity of the surface was uniform. Variations observed at the phase angle of 107° were about ± 50% of the extremely low value of 0.5%. Images of the Vega 2 flyby taken at low phase angles (about 20°) showed no identifiable variability of the surface reflectivity either. One has to conclude that the active areas (about 10% of the total surface as judged from the dust activity during the flyby, see below) were only insignificantly brighter than the inactive majority of the surface. Probably the interior of the nucleus was not physically different from its outer surface. There was no ice (in the classical sense) visible. The comet did not look differentiated.

The temperature of the dark surface (directly measured during the Vega 2 flyby, Emerich et al. (1987)) was so high that water ice could not exist on it. Therefore most of the nucleus was covered by a mantle. This surface mantle could well be a matrix of dust similar to the interior but simply depleted of water ice and other volatiles. It did not have to be a crust of regolith or debris left over from the sublimation activity. The thermal inertia of the surface layers seemed to be rather low, too low to keep activity going beyond the dusk terminator as witnessed by the dust distribution close to the nucleus (Thomas and Keller, 1988).

Topography

Many features on the surface manifested themselves not as much by variation of the surface reflectivity but by their imprints on the outline of the nucleus and of the terminator (Fig. 10). Only 25% of the visible surface were illuminated by the sun. The contributions of scattered light from dust in the very vicinity of the surface masked the light being scattered directly from the surface.

The most prominent feature directly visible was the 'crater', a roundish structure of about 2000 m in diameter, rather shallow 150 to 200

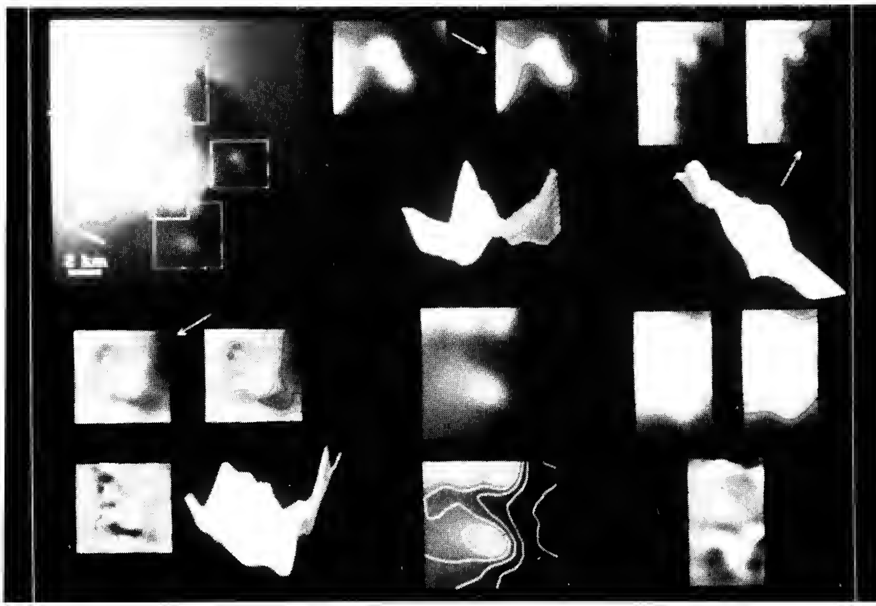


Fig. 10 - Details of the nucleus are shown in various cut outs that are individually stretched for contrast. Clockwise beginning at the middle of the top line: the «Mountain» is separated from the illuminated surface. Isopothes and 3-dimensional brightness contours are shown. The next cut out shows the «Finger» (north tip of nucleus) and the «Chain of Hills» undulating the terminator. Right hand lower corner: the area of activity at the central depression. The lower image is enhanced by unsharp masking revealing similar structures as on the northern part of the nucleus («Chain of Hills»). Lower line center: a small area of activity. Left hand side: the «Crater». The left most image is again enhanced by unsharp masking clearly revealing substructures of the order of 500 m within the «Crater».

m in depth (Schwarz et al., 1987). High resolution images showed structures of less than 400 m within the crater. Similar scales or slightly larges ones could be found in the undulation of the terminator north of the crater towards the northern tip of the nucleus. Here the strongest visible activity was located. Some of the very fine structures (filaments less than 500 m across) in the dust emission could be related to somewhat brighter spots on the surface. A very prominent feature was the bright spot within the dark section of the surface. The strong gradient in brightness was unresolved (less than 200 m). It could be the sun illu-

minimated tip of a 'mountain'. This interpretation implies a rise to abut 1 km above the ellipsoidal shape. A similarly accentuated feature was the 'duck tail', the southern rather sharp 90° corner of the nucleus seen in projection.

The coarse roughness of the nucleus implied a certain amount of tensile strength although gravity was very low. It also indicated that the sublimation did not equalize features. There were areas where the amount of volatiles within the surface region was depleted so that no sublimation could take place.

Rate of sublimation

It could be estimated that the loss of surface layer per revolution was about 6 m at active areas. For this estimate water sublimating unrestricted from a dark surface had been assumed and a density of the material of 1/3 g cm^3 (Huebner et al., 1986). If for example, the 'crater' was an area of enhanced and steady activity it would have been visible at the time of the first recorded apparition of comet Halley more than 2000 years (30 orbits) ago.

It would take about another several 1000 revolutions of comet Halley around the sun for its nucleus to vanish completely if the level of activity could be maintained. Investigations of meteor streams associated with the orbit of comet Halley show that the evolutionl of meteorite population has taken at least 2000 orbits of comet Halley around the sun on its present orbit (Hajduk, 1986). This makes comet Halley a rather old comet and indicates that the size of its nucleus was originally even larger, possibly twice its present linear dimension.

It also means that a possible alteration of the nuclear surface during the storage of the nucleus in the Oort cloud caused by cosmic rays over $4 \cdot 10^9$ *y* could not be found today. The thickness of such a layer is estimated to be substantially les than m (Johnson et al. 1987).

Rotation of nucleus

A period of rotation fo 52 to 54 h was derived from comparison of images taken during the three fly-bys (Wilhelm et al., 1986; Sagdeev et al., 1986b). Since it has not been possible to identify specific features on the Vega images a considerable uncertainty remains in the orientation of the spin axis (Keller and Thomas, 1988). The situation is further

complicated by nutation of the body. The variability of gas and dust production with a period of 7.3 or 14.6 d (Millis and Schleicher, 1986; Festou et al., 1987) is generally related to nutation (Sekanina, 1987; Julian, 1987). The rotation of about two days is also reflected in the curvature of dust jets.

The rather long rotation period (if compared to the mean period of asteroids) may be physically required for the nucleus to survive. Due to its elongated shape the velocity of rotation of the tips (8 km from the centre) is comparable to the orbital velocity around the nucleus if the average density of the nucleus is about 1/3 g cm^3.

Activity of the nucleus

One of the most unexpected results from the imaging was the strong concentration of dust jets emanating from limited areas only comprising about 10 to 20% of the total surface. This concentration of activity makes plausible the often observed rather strong variability of the cometary prodution within hours, short compared to the duration of a 'day'. It has not been possible to discern areas of activity clearly from those dormant.

Within these active areas (typical diameter about 3 km) a certain degree of inhomogeneity was visible expressed in the fine filaments of dust jet structures. These structures are strongly confined with opening angles of a few degrees. They could be caused by variation of the (visible) dust to gas ratio within the active area or by enhanced sublimation caused either by vatiation of physical properties or of chemical composition or indicate enhancement in the boundary layers of colliding jets.

Several faint dust jets (filaments) pointed in antisolar direction on images taken by HMC. They could still emanate from the area near the evening terminator. However, on Vega images at least one jet could be traced back to the dark side of the nucleus (Smith et al., 1986). Some places on the nucleus might have stored enough energy to support limited activity even on the night side.

Conclusions

The observations revealed the heart of a comet for the first time. The volatile component was dominanted by water as predicted. However, the appearance of the nucleus did not resemble a 'snow ball'. Its surfa-

ce was extremely dark and can be heated up to temperatures far beyond the sublimation temperature of water ice. The morphology and topography did not support a crust of left-over regolith of dust particles. The reflectivity varied little. The nucleus was not covered by a crust that is substantially different from the interior. Possible alterations due to high energy radiation have been lost on comets such as Halley.

The elongated and irregular shape of the cometary nucleus leads to the presumption that the nucleus was formed out of large subnuclei of kilometre size. This assumption is also supported by the typical scale size of surface structures in the range of 1/2 km. However smaller scale features were also observed as expected from a body size distribution.

The surface of the nucleus appeared uniform. Similarly, there was little variation within its active areas implying that the interior and the surface were of the same physical quality. No 'icy' surface was visible. The rather large topographic features (in particular the height of the mountain) eliminate the picture of a shrinking ice ball covered by regolith of larger dust particles. This all supports the picture of a nucleus the physical structure of which is dominated by the matrix of the nonvolatile (dust) rather than by the volatile material (ice). Large parts of the surface were depleted of volatiles and could reach high temperatures. The depth of this layer of depletion is unknown but does not have to be thick. The surface texture was probably very fluffy as witnessed by the observed dust particles. This explained the extremely low reflectivity. The heat conduction by the solid material will be low. However, heat could be transported by the sublimated gas inside the fluffy surface which could recondense further inside.

The scenario of the formation of cometary nuclei could be as follows. Ice covered dust particles (not only silicates but rather condensates of nonvolatile (compared to water) material including organic substances) form fluffy structures. The ice is predominantly preserved within the cavities. Cometary nuclei may may have been formed out of dust particles rather than of 'snow flakes'. The ice filled dust particles clump together to form larger grains and bodies. Collisions even at low relative speeds lead to strong perturbations of the contact areas. The bodies are so fluffy (but still stiff) that they can penetrate each other during the collisions. These contact zones may show up as ihomogeneities leading to changes in the rate of sublimation. The requirement of noncatastrophic collisions restrict the formation of the nuclei to a region with low relative velocities near the limits of the present planetary system. Not only the sizes of the nuclei but also their masses ($\approx 10^{17}$ g) are rather large.

REFERENCES

A'Hearn, M., Hoban, S., Birch, P.V., Bowers, C., Martin, R., Klinglesmith, D.A.: 1986, *Nature* **324**, 649.

Balsinger, H., Altwegg, K., Bühler, F., Geiss, J., Ghielmetti, A.G., Goldstein, B.E., Goldstein, R., Huntress, W.T., Ip, W.H., Lazarus, A.J., Meier, A., Neugebauer, M., Rettenmund, U., Rosenbauer, H., Schwenn, R., Sharp, R.D., Shelley, E.G., Uungstrup, E., Young, D.T.: 1986, *Nature* **321**, 330.

Clark, B.C., Mason, L.W., Kissel, J.: 1987, *Astron. Astrophys.* **187**, 779.

Crifo, J.F.: 1987, *Xth European Regional Astronomy meeting of the I.A.U.*

Curdt, W., Wilhelm, K., Craubner, H., Krahn, E., Keller, H.U.: 1988, *Astron. Astrophys.* **191**, L1.

Eberhardt, P., Krankowsky, D., Schulte, W., Dolder, U., Lämmerzahl, P., Berthelier, J.J., Woweries, J., Stubbemann, U., Hodges, R.R. Hoffman, J.H., Iliano, J.M.: 1986, *Symposium on the Exploration of Halley's Comet* **ESA-SP 250**, 383.

Eberhardt, P., Krankowsky, D., Schulte, W., Dolder, U., Lämmerzahl, P., Berthelier, J.J., Woweries, J., Stubbemann, U., Hodges, R.R., Hoffmann, J.H., Illiano, J.M.: 1987, *Astron. Astrophys.* **187**, 481.

Emerich, C., Lamarre, J.M., Moroz, V.I., Combes, M., Sanko, N.F., Nikolsky, Y.V., Rocard, F., Gispert, R., Coron, N., Bibring, J.P., Encrenaz, T., Crovisier., J.: 1987, *Astron. Astrophys.* **187**, 839.

Festou, M.C., Drossart, P., Lecacheux, J., Encrenaz, T., Puel, F., Kohl-Moreira, J.L.: 1987, *Astron. Astrophys.* **187**, 575.

Galeev, A.A.: 1987, *Astron. Astrophys.* **187**, 12.

Gringauz, K.I., Gombosi, T.I., Remizov, A.P., Apáthy, I., Szemerey, I., Verigin, M.I., Denchikova, L.I., Dyachkov, A.V., Keppler, E., Klimenko, I.N., Richter, A.K., Somogyi, A.J., Szegö, K., Szendrö, S., Tátrallyay, M., Varga, A., Vladimirova: 1986, *Nature* **321**, 282.

Haiduk, A.: 1986, *Symposium on the Exploration of Hally's Comet* **ESA-SP 250**, 239.

Huebner, W.F.: 1987, *Science* **237**, 628.

HUEBNER, W.F., DELAMERE, W.A., REITSEMA, H.J., KELLER, H.U. WILHELM, K., WHIPPLE, F.L., SCHMIDT, H.U.: 1986, *Symposium on the Exploration of Halley's Comet* **ESA-SP 250**, 363.

HUEBNER, W.F., BOICE, D.C., SHARP, C.M., KORTH A., LIN, R.P., MITCHELL, D.L., RÈME, H.: 1987, *Symposium on the Diversity and Similarity of Comets* **ESA-SP 278**, 163.

IP, W.-H.: 1988, On Charge Exchange Effect in the Vicinity of the Cometopause of comet Halley, *preprint.*

JESSBERGER, E.K., CHRISTOFORIDIS, A., KISSEL, J.: 1988, *Nature* **332**, 691.

JOHNSON, R.E., COOPER, J.F., LANZEROTTI, L.J., STRAZZULA, G.: 1987, *Astron. Astrophys.* **187**, 889.

JOHNSTON, A., COATES, A., KELLOCK, S., WILKEN, B., JOCKERS, K. ROSEMBAUER, H., STÜDEMANN, W., WEISS, W., FRMISANO, V., AMATA, E., CERULLI-IIRELLI, R., DOBROWOLNY, M., TERENZI, R. EGIDI, A., BORG, H., HULTQUIST, B., WINNINGHAM, J., GURGIOLO, C., BRYANT, D., EDWARDS, T., FELDMAN, W., THOMSEN, M., WALLIS, M.K., BIERMANN, L., SCHMIDT, H., LUST, R., HAERENDEL, G., PASCHMANN, G.: 1986, *Nature* **321**, 344.

JULIAN, W.M.: 1987, *Nature* **326**, 57.

KANEDA, E., HIRAO, K., TAKAGI, M., ASHIHARA, O., ITHO, T., SHIMIZU, M.: 1986a, *Nature* **321**, 140.

KANEDA, E., ASHIHARA, O., SHIMIZU, M., TAKAGI, M., HIRAO, K.: 1986b, *Nature* **321**, 297.

KELLER, H.U., THOMAS, N.: 1988, *Nature* **333**, 146.

KELLER, H.U., ARPIGNY, CLAUDE, BARBIERI, C., BONNET, C.M. CAZES, S., CORADINI, M., COSMOVICI, C.B., DELAMERE, W.A., HUEBNER, W.F., HUGHES, D.W., JAMAR, C., MALAISE, D., REITSEMA, H.J., SCHMIDT, H.U., SCHMIDT, W.K.H., SEIGE, P., WHIPPLE, F.L. WILHELM, K.: 1986, *Nature* **321**, 320.

KELLER, H.U., SCHMIDT, W.K.H., WILHELM, K., BECKER, CH., CURDT, W., ENGELHARDT, W., HARTWIG, H. KRAMM, J.R., MEYER, H.J., SCHMIDT, R., GLIEM, F., KRAHN, E., SCHMIDT, H.P., SCHWARZ, G., TURNER, J.J., BOYRIES, P., CAZES, S., ANGRILLI, F., BIANCHINI, G., FANTI, G., BRUNELLO, P., DELAMERE, W.A., REITSEMA, H.J., JAMAR, C., CUCCIARO, C.: 1987a, *J. Phys. E.: Sci. Instrum.* **20**, 807.

KELLER, H.U., DELAMERE, W.A., HUEBNER, W.F., REITSEMA, H.J., SCHMIDT, H.U., WHIPPLE, F.L., WILHELM, K., CURDT, W., KRAMM, J.R., THOMAS, N., ARPIGNY, CLAUDE, BARBIERI, C., BONNET, R.M., CAZES, S., CORADINI, M., COSMOVICI, C.B., HUGHES, DAVID, W., JAMAR, C., MALAISE, D., SCHMIDT, K., SCHMIDT, W.K.H., SEIGE, P.: 1987b, *Astron. Astrophys.* **187**, 807-823.

KISSEL, J., SAGDEEV, R.Z., BERTAUX, J.L., ANGAROV, V.N., AUDOUZE, J., BLAMONT, J.E., BÜCHLER, K., EVLANOV, E.N., FETCHTIG, H., FOMENKOVA, M.N., von HOERNER, H., INOGAMOV, N.A., KHROMOV, V.N., KNABE, W., KRUEGER, F.R., LANGEVIN, Y., LEONAS, V.B., LEVASSEUR-REGOURD, A.C., MANAGADZE, G.G., PODKOLZIN, S.N., SHAPIRO, V.D., TABALDYEV, S.R., ZUBKOV, B.V.: 1986a, *Nature* **321**, 280.

KISSEL, J., BROWNLEE, D.E., BÜCHLER, K., CLARK, B.C., FECHTIG H., GRÜN, E., HORNUNG, K., IGEBERGS, E.B., JESSBERGER, E.K., KRUEGER, F.R., KUCZERA, H., MCDONNEL, J.A.M., MORFILL, G.M., RAHE, J., SCHWEHM, G.H., SEKANINA, Z., UTTERBACK, N.G., VÖLK, H.J., ZOOK, H.A.: 1986b, *Nature* **321**,336.

KORTH, A., RICHTER, A.K., LOIDL, A., ANDERSON, K.A., CARLSON, C.W., CURTIS, D.W., LIN, R.P. RÈME, H., SAUVAUD, J.A., D'USTON, C., COTIN, CROS, A., MENDIS, D.A.: 1986, Nature **321**, 335.

KORTH, A., MARCONI, M.L., MENDIS, D.A., KRUEGER, F.R., RICHTER, A.K., LIN, R.P., MITCHELL, D.L., ANDERSON, K.A., CARLSON, C.W., RÈME, H., SAUVAUD, J.A., D'USTON C.: 1988, *Nature, in press.*

KRANKOWSKY, D., EBERHARDT, P.: 1988, in: *Comet Halley 1986: World-Wide Investigations, Results and Interpretations.* Ed. J. Mason, P. Moore, Chichester: Ellis Horwood, in press.

KRANKOWSKY, D., LÄMMERZAHL, P., HERRWERTH, I., WOWERIES, J., EBERHARD, P., DOLDER, U., HERRMANN, U., SCHULTE, W., BERTHELIER, J.J., ILIIANO, J.M., HODGES, R.R., HOFFMAN, J.H.: 1986, *Nature* **321**, 326.

MCDONNELL, J.A.M., AIEXANDER, W.M., BURTON, W.M., BUSSOLETTI, E., EVANS, G.C., EVANS, S.T., FIRTH, J.G., GRARD, R.J.L., GREEN, S.F., GRUN, E., HANNER, M.S., HUGHES, D.W., IGENBERGS, E., KISSEL, J., KUCZERA, H., LINDBLAD, B.A., LANGEVIN, Y., MANDEVILLE, J.-C., NAPPO, S., PANKIEWICZ, G.S.A., PERRY, C.H., SCHWEHM, G.H., SEKANINA, Z., STEVENSON, T.J., TURNER, R.F., WEISHAUPT, U., WALLIS, M.K. ZARNECKI, J.C.: 1987, *Astron. Astrophys.* **187**, 719.

MILLIS, R.L., SCHLEICHER, D.G.: 1986, *Nature* **324**, 646.

MUKAI, T., MIYAKE, W., TERASAWA, T., KITAYAMA, M., HIRAO, K.: 1986, *Nature* **321**, 299.

NEUBAUER, F.M., GLASSMEIER, K.H., POHL, M., RAEDER, J., ACUNA, M.H., BURLAGA, L.F., NESS, N.F., MUSMANN, G., MRIANI, F., WALLIS, M.K., UNGSTRUP, E., SCHMIDT, H.U.: 1986, *Nature* **321**, 352.

SAGDEEV, R.Z., SZABÓ, F., AVANESOV, G.A., CRUVELLIER, P., SZABÓ, L., SZEGÖ, K., ABERGEL, A., BALAZS, A., BARINOV, I.V., BERTAUX, J.-L., BLAMONT, J., DETAILLE, M., DEMARELIS, E., DUL'NEV, G.M., ENDRÖCZY, G., GARDOS, M., KANYO, M., KOSTENKO, V.I., KRASIKOV, V.A., NGUYEN-TRONG, T., NYTRAI, Z., RENY, I., RUSZNYAK, P., SHAMIS, V.A., SMITH, B., SUKHANOV, K.G., SZABÓ, F., SZALAI, S., TARNOPOLSKY, V.I., TOTH, I., TSUKANOVA, G., VALNICEK, B.I., VARHALMI, L., ZAIKO, YU. K., ZATSEPIN, S.I., ZIMAN, YA.L., ZSENEI, M., ZHUKOV, B.S.: 1986a, *Nature* **321**, 262.

SAGDEEV, R.Z., KRASIKOV, V.A., SHAMIS, V.A., TARNOPOLSKI, V.I., SZEGÖ, K., TÓTH, I., SMITH, B., LARSON, S., MERENYI, E.: 1986b, *Symposium on the Exploration of Halley's Comet* **ESA-SP 250**, 335.

SCHWARZ, G., CRAUBNER, H., DELAMERE, W.A., GOEBEL, M., GONANO, M., HUEBNER, W.F., KELLER, H.U., KRAMM, J.R., MIKUSCH, E., RREITSEMA, H.J., WHIPPLE, F.L., WILHELM, K.: 1987, *Astron. Astrophys.* **187**, 847.

SEKANINA, Z.: 1987, *Nature* **325**, 326.

SEKANINA, Z.: 1988, *Outgassing Asymmetry of Periodic Comet Encke. II Apparitions 1868-1918 and a Study of the Nucleus Evolution*, submitted to Astron. Astrophys.

SEKANINA, Z., LARSON, S.M.: 1986, *Astron. J.* **92**, 462.

SMITH, B., SZEGÖ, K., LARSON, S., MERÉNYI, E., TÓTH, I., SAGDEEV, R.Z., AVANESOV, G.A., KRASIKOV, V.A., SHAMIS, V.A., TARNAPOLSKY, V.I.: 1986, *Symposium on the Exploration of Halley's Comet* **ESA-SP 250**, 327.

THOMAS, N., KELLER, H.U.: 1988, *Proceedings of the Symposium Dust in the Universe, in press.*

WHIPPLE, F.L.: 1986, *Symposium on the Exploration of Halley's Comet* **ESA-SP 250**, 281.

WILHELM, K., COSMOVICI, C.B., DELAMERE, A.W., HUEBNER, W.F., KELLER, H.U., REITSEMA, H.J., SCHMIDT, H.U., WHIPPLE, F.L.: 1986, *Symposium on the Exploration of Halley's Comet* **ESA-SP 250**, 367.

Celestial Mechanics
and space-related activities

Jean KOVALEVSKY *

I - INTRODUCTION

I have been requested by Professor Romano to present this review on the interaction and cross fertilization between Celestial Mechanics and Astronautics. This subjet is in the same time quite vast and also rather vague in the sense that the frontier between Celestial Mechanics and Astronautics is not clearly defined. Just to take an example, a new word has been introduced in the early sixties: «astrodynamics«. This was a very confusing name since it meant dynamics (which is a little more restrictive than mechanics) applied to celestial bodies — astro — but meant actually mainly artificial celestial bodies. So one could identify Astrodynamics to a section of Celestial Mechanics. On the other hand, the implicit reference to artificial celestial bodies would authorize also to consider that Astrodynamics is part of Astronautics. For this reason, I prefer to use the more general wording of space-related activities and let people decide for themselves what belongs to what category.

Astronautics were born 30 years ago under the double auspices of rocket engineering which permitted to put objects in space and Celestial Mechanics which told how the object will move in space. But Celestial Mechanics itself has greatly evolved since then and modern Celestial Mechanics has also two parents: the old-time Celestial Mechanics and Computer Sciences. So the relationships that one may find between Astronautics and Celestial Mechanics are twofold: first, what was the initial input of Celestial Mechanics into the newly born space sciences and secondly, what have been, and still are the connections between the fastly evolving modern Celestial Mechanics and space activities. So, let us first recall what was the state of Celestial Mechanics and what part of it was useful to astronautics.

* CERGA (France).

II - CELESTIAL MECHANICS IN THE LATE FIFTIES

Before the launch of the first artificial satellite, Celestial Mechanics was not a very active field as a whole. However a limited number of very important studies were pursued then, leading the way to the expansion of the next decades. Let us sketch some of these fields.

1. Theoretical Celestial Mechanics

On the steps of Poincaré, Birkhoff and Levi Civita, the school of Kustaanheimo, Stiefel, Siegel, etc. ... developed various theories of regulatization that proved to be very useful later in numerical work. But the major break-through was the famous theorem of Arnold-Kholmogorov-Moser which was the origin of a large number of developments in the understanding of the structure of the manifold of solutions of dynamical problems.

2. Planetary and satellite theories

The last improvements of a series of analytical theories initiated in the XIX-th century by Leverrier and Newcomb were being implemented by G.M. Clemence for Mars (1949 and 1961) and W.T. Eckert for the Moon (Eckert and Smith, 1966). They introduced the use of litteral algebra by computers. Similarly, computer techniques in semi-numerical theories were initiated for satellite theory by Kovalevsky (1959). It is also interesting to mention the theory of the V-th satellite of Jupiter by A.J.J. Van Woerkom (1950), since it was a preview of the theory of the motion of an artificial satellite.

3. Numerical integration of a few body problem

Probably the newest and the most promising for the future undertaking of that decade was the numerical integration of the motion of the four outer planets over 400 years by W.J. Eckert, D. Brouwer and G.M. Clemence (1951). It has been the model for many other numerical integrations in the future and popularized the predictor-corrector method in the form devised by Cowell for Celestial Mechanics.

This was the state of the art for advanced Celestial Mechanics at the eve of the space era, but of course the few Celestial Mechanists in the world held the deposit of two centuries of capitalized methods, techniques and results in the treatment of the equations of motion of celestial

objects. It is in this treasury rather than in the foremost advanced activities of research in Celestial Mechanics that the newly born Astronautics took the major part of knowledge that permitted its developments.

III - THE FIRST YEARS OF SPACE AGE

The launch of the first artificial satellite provided Celestial Mechanicists with a remarkable problem: the motion of a satellite around a planet. Let us remind that it consists in finding the trajectories of a massless body in a field defined by the following potential:

$$U = \frac{\mu}{r} - \frac{\mu R^2}{r^3} J_2 P_2 (\sin \phi) \tag{1a}$$

$$- \sum_{n=3}^{\infty} \frac{\mu R^n}{r^{n+1}} J_n P_n (\sin \phi) \tag{1b}$$

$$+ \sum_{n=2}^{\infty} \sum_{k=1}^{n} \frac{\mu J_{nk} R^n}{r^{n+1}} P_{nk} (\sin \phi) \cos (\lambda - \lambda_{nk}) \tag{1c}$$

where r, λ and ϕ are the equatorial geocentric spherical coordinates of the body, and where one has the notations:

μ: geocentric constant of gravitation
R: equatorial radius of the Earth
J_2, J_n: zonal harmonics
J_{nk}, λ_{nk}: tesseral harmonics
P_n and P_{nk}: Legendre polynomials and associate Legendre functions.

This problem had all the qualities to stir up the interest of Celestial mechanicists and to attract many new people in the field.

1. It was a useful problem. It was necessary that it be solved and the solution be applied to the new astronautics.
2. It was much simpler that the classical planetary or satellite theories so that many beginners could cope with it. But it could become complicated at will simply by adding new terms in the Earth potential.

These two reasons attracted a large number of people into Celestial

Mechanics and the part of astronautics that dealt with the motion of artificial celestial bodies. May I mention just one revealing fact: while D. Brouwer, who was certainly the leading professor of Celestial Mechanics in United States, used to have one or two students or simply listeners to his course of Celestial Mechanics at Yale University, his «Summer Institute for Dynamical Astronomy» attracted from the first session, more than sixty students.

The artificial satellite problem played indeed an enormous role. Its simplest form, the J_2 term of the potential (part 1a of the formula), is a remarkable two degrees of freedom problem. It led to several major developments.

— Brouwer (1959) solved it by introducing the so-called Von Zeipel method - an improvement of the Poincaré-Lindstedt method. He also, with several others, pointed out the singularity at the critical inclination that was studied for many years as a quasi-ideal case of the simple resonance.
— Several scientists have applied independently and successfully to it the Lie series a method that also proved to be powerful.
— Several other methods, using Lagrange or Gauss equations were also applied by many other people.
— The same potential is very closely represented by the potential of two equal masses separated by an imaginary distance. It corresponds to (1a) plus (1b) where all the zonal harmonics are zero except $J_4 = J_2^2$ .This special two point problem studied by Vinti (1961), is an integrable problem that was extensively used in particular by the Moscow school (Akcionov, Grebenikov, Diomin) as the basis of more complete theories of the motion of artificial satellites.

It is to be noted that the solution by Van Woerkom ten years before of a similar problem for the V-th satellite of Jupiter did not stir up any interest. This shows the impact of the space age on Celestial Mechanics and the importance of the increase of the number of people involved.

In parallel to or as a consequence of the first studies on the restricted potential, second order solutions in J_2 were constructed as well as analytical solutions at any order of even and odd zonal harmonics (part 1a and 1b of the potential). These were applied for the determination of the corresponding parameters of the Earth's potential from secular and long periodic terms in the theory of the motion of various satellites by a great number of scientists of different countries. The tesseral harmonic theory was also constructed and used under its analytical form to produce the first «Standard Earth» (Lundquist and Veis, 1966). Other

attemps to use analytical methods were successfully made for atmospheric drag (Brouwer and Hori, 1961), radiation pressure, etc...

It is also from these first studies that some interesting features of artificial satellite motion were identified and led to the conception of many space systems: heliostationary satellites for meteorology, geostationary stable orbits, undisturbed configuration of polar satellites for navigation, recurrent orbits for oceanographic studies, stabilization by gravity gradient, tethered satellites, etc...

But, with time, the requested accuracy for space navigation or orbit determination increased and became unaccessible to existing analytical solutions. They were replaced by numerical integration techniques initiated in Celestial mechanics by Brouwer et al. With the advent of faster and more powerful computers, this techniques made remarkable progress and became the basic method for practically all orbit engineering.

IV - SPECIFIC PROBLEMS OF ASTRONAUTICS

The tremendous appeal to Celestial mechanics that marked the early sixties when the problems described in the preceding section were of interest to Celestial Mechanicists for their scientific impact and the newly born Astronautics for their applications ceased several years later and the number of papers on this subject decreased drastically, mostly because the problems were solved to the satisfaction of both sides. Some of the people who worked in this field remained in Celestial Mechanics shifting their interest to more classical domains of this science. other applied Celestial Mechanics to other scientific problems opened by Space Research or applied its methods in analysing data from space experiments (see next section). A third group developed methods for engineering or mission analysis problems specific to astronautics, methods that were no more based on Celestial Mechanics alone. Let us describe some of the problems involved.

1. Optimization of orbit transfer

The problem is to find a trajectory - and in a first approximation one or several successive Keplerian orbits - that permits to transfer from a given orbit to another one.

The optimization is usually, in a first approximation, made assuming impulsive changes of state, which is a good approximation to the action of thrusters. The optimization may search for minimum fuel trajectory

(equivalent to a minimum velocity increment trajectories) or minimum time of transfer, or a given weighted combination of them. In the simplest cases, one has to apply the variational calculus and the theory of optimization. Actually, one or several keplerian transfer orbits can permit to arrive at a close enough solution so that the variational methods can be later applied through numerical integration techniques. Among the simple transfer problems, one may list:
— Hohman transfer between two circular coplanar orbits,
— Bielliptic transfers using two elliptic coplanar orbits,
— Non coplanar transfers,
— Optimum intercept maneuvres.

It is of course applied to rendez-vous, but also to lunar transfers or to interplanetary missions.

2. Slow orbital maneuvres

A very interesting problem - not yet applied in practice, but with a very promising future - is the possibility of using a continous very weak force for space navigation. This is the case of future ion motors or of the solar sail. This problem matches very well with methods of Celestial Mechanics, because it can be treated as a perturbation problem, at least for a short time. But the maneuvres, that is to say the control of the thrusters, is no more akin to Celestial Mechanics and is a real problem of astronautics with its own methods.

3. Maneuvres using planetary approaches

What happens to a comet when it approaches a planet is an important problem of Celestial mechanics. It was first solved by Tisserand (1896) with his famous criterion using the Jacobi integral:

$$\frac{1}{a_0} + \frac{2\sqrt{a'}\sqrt{p_0}\cos i_0}{r'^2} = \frac{1}{a_1} + \frac{2\sqrt{a'}\sqrt{p_1}\cos i_1}{r'^2}$$

where r' is the planetary mean radius and a' its semi-major axis:

$$p_0 = a_0\,(1-e_0^2) \qquad\qquad p_1 = a_1(1-e_1^2)$$

i_0 and i_1 are inclinations with respect to the orbital plane of Jupiter before and after the encounter.

A somewhat more exact method uses the concept of sphere of influence (Laplace):

$$R_{\text{infl}} = r_0 \left(\frac{\mathrm{m}_1}{m_3} \right)^{2/5}$$

r_0 = distance planet-Sun
m_1 = mass of the planet
m_3 = mass of Sun.

Inside the radius R_{infl}, the motion is planetocentric, outside, heliocentric, and the trajectory is represented in a first approximation by a segmented trajectory consisting in patched conic segments. The finalization of the trajectory has of course to be done by numerical integration and by optimization techniques if one wishes to reach a given orbit. This technique was used to achieve some of the most spectacular space mission.

VOYAGER 2: A triple swing - by maneuvres successively with Jupiter, Saturn and Uranus permitting close approaches to Saturn, Uranus and Neptune as shown in figure 1.

VEGA: This space mission visited first Venus and swing - by maneuvres directed it to a close approach to comet Halley.

ISEE 3: This satellite was displaced from its position near a libration point and, using several swing - by maneuvres near the Moon, was put on an heliocentric orbit in order to pass close to the Giacobini - Zinner comet.

ULYSSES: The heliocentric orbit perpendicular to the ecliptic will be obtained by a fly-by near Jupiter.

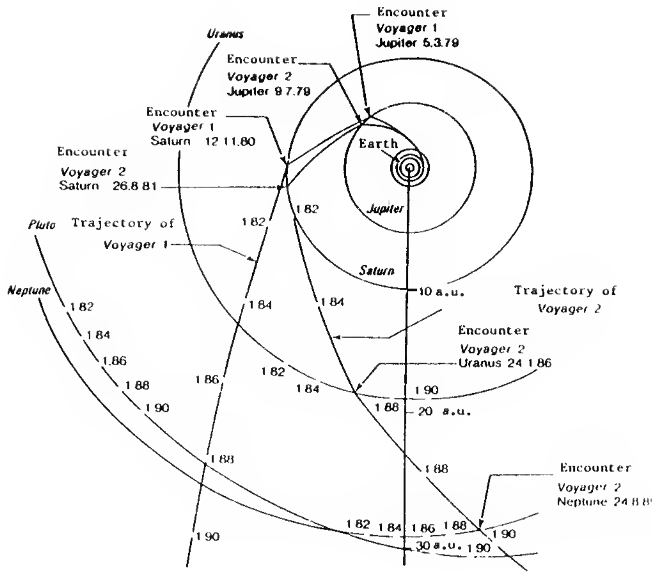


Figure 1 - Trajectories of VOYAGER 1 and 2 in their missions to the outer planets.

4. Rotation of a spacecraft

The problem of the attitude of a satellite and of a spacecraft is fundamental in astronautics, either for controlling it or for recovering it. A satellite for which the attitude is lost may often be a lost satellite. Celestial Mechanics have not contributed very much to the problem, even if the rotation of celestial bodies was studied for the Moon, the Earth and more recently for Venus and Mercury. But these were very particular problems, while the general case of a rotation of a solid body has not attracted Celestial Mechanicists. The lunar case was a prefiguration of the gravity gradient attitude control but that was very elementary in comparison with the needs of astronautics. Actually, this subject was mainly driven by aeronautics in its engineering aspects (Tait Bryan system of angular coordinates: roll, pitch and yaw angles) and was derived from classical mechanics for more scientific aspects (Belyaev and the USSR school, and USA University Spacecraft Dynamics researches) for the construction of mathematical models of the gravitating bodies and then for the solution.

So the stabilization problems (gravity gradient, magnetic, inertial wheels and gas jets) as well as the pointing problem (which is an optimization problem of transfer between two attitudes) or attitude reconstitution are specific astronautical problems solved without great reference to Celestial Mechanics.

Of course there are exceptions. For instance, in the ESA project HIPPARCOS, a very accurate knowledge of the attitude (0.″1 in 2 directions, a few milliseconds of arc on the rotation proper) was necessary for the obtention of the scientific results. Of course the description of the attitude could always be done numerically, either by numerical integration or by numerical representation by an adequate set of functions. However, recently in a thesis, E. Bois (1986) has shown that perturbation theory of Celestial Mechanics may apply to a good accuracy to such a problem. Figure 2 shows an example of variations of attitude angles of HIPPARCOS, controlled by a closed loop system acting on gas jet actuations. They can be represented by trigonometric series of time with a linear function of time changing at each gas jet actuations (Belforte et al., 1983).

5. Dynamics of the solar system

Navigation in the solar system necessitates a very precise knowledge of the positions and masses of all objects of the solar system. So astronautics required a very great effort in improving the ephemerides of all planets. This was a challenge to Astronomy and Celestial Mechanics, that was met using classical astronomical observations of bodies in the solar system, radar distances, lunar laser ranging, precise trajectography of spacecrafts with Doppler data, and VLBI observations while they were near the planets as the Vikings around Mars or lunar ALSEP. The most important effort is the Ephemeris development program in JPL which lasts 20 years during with a continuous refinement and analysis of data reduction process.

The present accuracy achieved is, in a dynamical self consistent frame of reference:

object	position	mean motion per century
Mercury	0.″03	0.″15
Venus/Earth/Mars	0.″01	0.″03 to 0.″06
Outer planets	0.″30	0.″7 to 1″
Moon	0.″01	0.″65 per century squared

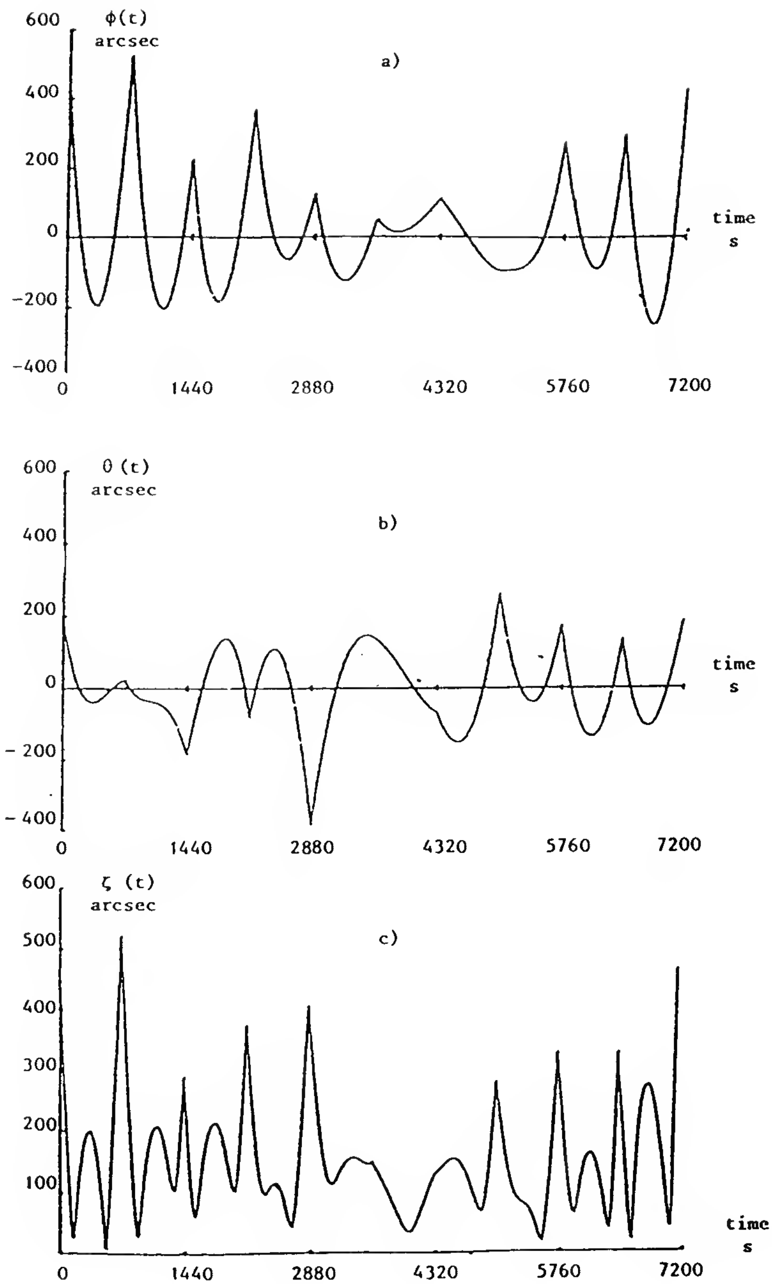


Figure 2 - Modelled attitude of HIPPARCOS during one revolution. The angles represent the differences with respect to nominal attitude variation.

It is expected that the final reduction of Voyager 2 data would improve by a factor of 2 or 3 the accuracies for the outer planets. It is also to be remarked that these numbers refer to a consistent dynamical system of reference. If one was to use the FK4 system, the uncertainties for the Moon and the terrestrial planets would be degraded by a factor of 3 to 5 (Standish, 1986).

In some way, the major input of Celestial Mechanics and Dynamical astronomy to Astronautics are the accurate ephemerides in the solar system for which gains of a factor 100 for the Moon, 20 for the inner planets, 5 still to be improved for the outer planets have been achieved.

V - CELESTIAL MECHANICS IN SPACE RESEARCH

Conversely, Celestial Mechanics has immensely gained from space research or, we should better say, that Celestial Mechanics was a major tool for interpreting space data and discoveries made by space missions have provided Celestial Mechanicists with new and very exciting problems. Let us quote some, but the list is very far from being exhaustive.

1. Artificial satellite motion

We have already stated that the first Earth gravity model was constructed using only analytical Celestial Mechanics. The trend has amplified since then, even though numerical methods have been used. The results from the dynamics of satellite motion have been very numerous in a wide range of domains.

- The Earth potential models are now expressed with more than 1000 different terms (500 x 500 km resolution).
- The models of the upper atmosphere and thermosphere are based upon the analysis of orbital changes of satellites undergoing air drag.
- Very detailed studies of the dynamics of some satellites such as D5B, STARLETTE, LAGEOS, etc... yielded results on:
  - the tidal deformations of the Earth potential for which a model with several parameters was obtained,
  - the effects of radiation pressure direct or diffused by the Earth as well as infrared Earth radiation using accelerometers on D5B,
  - the photo-ionic reaction due to differences of temperature between opposite parts of the surface.
- The observation of the motion of artificial satellites by laser ranging in comparison with the theory of their motion permits to obtain one

of most precise determinations of the Earth rotation parameters.
— Using all the results, the motion of LAGEOS is now modelled to a fraetion of a centimeter.

2. Planetary gravity fields

— The lunar probes have been a very efficient tool to determine the gravity field of the moon. The discovery of so-called mascons - regions of higher gravity field that can be modelled by underground mass concentrations - have been one of the major achievements of the lunar orbiter programs.
— The observations of the Viking satellites around Mars have been used to determine the gravity field of the Moon up to the tesseral harmonics of order 18. the gravity anomaly produced by the huge vulcano Olympic Mons is very conspicuous in these results.
— Some results on the gravity field of Venus have also been obtained.

3. Dynamics of the solar system

The tracking data of various space-probes to Venus, Mars, Mercury and the outer planets gave observational conditions to the theory of the motion of the planets in the solar system that yielded significant improvement in the parameters of their motion and in the planetary masses.
In particular, about one order of magnitude could be gained in the evaluation of the parameters of the General Relativity.

4. New problems in Celestial Mechanics

New data acquired by space research in the solar system raised new important problems in Celestial Mechanics that were not suspected. The most striking examples to which many people including Giuseppe Colombo have contributed have been raised by the situations discovered in the Saturn system by the two Voyager missions, although the discovery from the ground of the strange elliptic shape of the Uranian rings was a first recongnition of a new type of gravitational effects. These findings include:
— The dynamics of coorbiting satellites that never collide, through the interexchange of orbits.
— The containment of rings by two shepherd satellites outside and inside the ring.
— The thin ringlets structure of the Saturn ring system due probably to

a similar effect of reciprocal containment of ringlet material, a ringlet being in some way a gravitational fence for the adjacent rings.

But these are not the only examples. Many other problems of Celestial Mechanics sprung from the results of space research or using data that was collected by means of various space vehicules. Let us just quote some of them.

— The evaluation of the lunar free libration from lunar laser ranging results.
— The confirmation to a better accuracy of the principle of equivalence in the Earth-Moon system using also lunar-laser ranging observations.
— The extension to satellite systems (in particular Saturn system) of the trojan stable regions.
— The studies related to slow resonant rotation of Mercury and Venus confirmed by the various space probes sent to these planets.
— The very refined determination of the orbit of Halley comet partly from the early probe observations in order to prepare the Giotto encounter.
— The study of new families of periodic orbits in conjunction with the search or the realization of swing-by maneuvres.
— The introduction of the general theory of the motion of the planets in the solar system in the formalism of the General Relativity theory rather than considering a Newtonian field with relativistic perturbations.
— etc...

VI - CONCLUSION

While the dynamical problems of astronautics and their solutions are partly issued from Celestial Mechanics, conversely Space Research has provided celestial Mechanicists many important and fascinating domains of study. This cross-fertilization has indeed been very profitable for both sciences and many scientists have been actively involved in this double domain. Among them one of the most proeminent was certainly Giuseppe Colombo.

BIBLIOGRAPHY

N.B. To give even a partial account of the litterature published in the domains covered by this lecture would be an immense task with unavoidable omissions and unfairness. This is why we have identified only those references that are representative of the status of Celestial mechanics at the beginning of the Space Age and the few papers that have opened the way to the first achievements in artificial satellite theory. I have only added to this stet references that illustrate a couple of points that are not widely known and that I could not develop here in details.

BELFORTE, P., CANUTO, E., DONATI, F. and VILLA, A., 1983, in «The FAST think-shop», P.L. Bernacca ed., University of Padova, p. 163.

BOIS, E., 1986, Celestial Mechanics, vol. 39, p. 309.

BROUWER, D., 1959, Astron. Journal, vol. 64, p. 378.

BROUWER, D. and HORI, G., 1961, Astron. Journal, vol. 66, p. 193.

CLEMENCE, G.M., 1949, Astron. Papers of the Amer. Ephemeris, vol. XI, part. 2.

CLEMENCE, G.M., 1961, Astron. Papers of the Amer. Ephemeris, vol. XVI, part. 2.

ECKERT, W.J., BROUWER, S. and CLEMENCE, G.M. 1951, Astron. Papers of the Amer. Ephemeris, vol. XII.

ECKERT, W.J. and SMITH, H.F. Jr., 1966, Astron. Papers of the Amer. Ephemeris, vol. XIX, part. 2.

KOVALEVSKY, J., 1959, Bulletin astronomique, vol. 23, p. 1.

LUNDQUIST, C.A. and VEIS, G., 1966, Smithsonian Astrophysical Observatory Special Report, n° 200.

STANDISH, E.M., 1986, in «Relativity in Celestial Mechanics and Astrometry», J. Ko valevsky and V.A. Brumberg eds., D. Reidel Publ., p. 71.

TISSERAND, G., 1896, «Traité de Mécanique Celeste», Gauthier-Villars, Paris, vol. 4, p. 203.

VAN WOERKOM, A.J.J., 1950, Astron. Papers of the Amer. Ephemeris, vol. XIII, part. 1.

VINTI, J.P., 1961, Astron. Journal, vol. 64, p. 514.

Gravitazione e modelli di spazio-tempo

Leonid I. SEDOV

Lo studio e la descrizione di fenomeni fisici sono legati all'introduzione e all'utilizzo di sistemi di riferimento e di modelli metrici e geometrici dello spazio-tempo.

Nelle applicazioni vengono diffusamente impiegati concetti legati al tempo assoluto, nello spazio tridimensionale euclideo e negli spazi tetradimensionali di Riemann nella teoria della relatività ristretta e in quella generale.

Il presente lavoro è dedicato alla teoria generale della gravitazione, inscindibile dalle teorie sui modelli tetradimensionali di spazi pseudoriemanniani.

È possibile introdurre in questi spazi delle coordinate e osservare le famiglie di linee di universo K, corrispondenti al singolo movimento di particelle sperimentali e materiali «di pulviscolo» (materia disgregata), non soggette a collisioni; il movimento può essere considerato come un continuo puntiforme descritto nelle coordinate lagrangiane (coordinate comobili) $\tau = \xi^4$, ξ^α $(\alpha = 1, 2, 3)$, dove ogni linea della famiglia K corrisponde a un valore delle coordinate ξ^α . Nei punti della famiglia di linee K è sempre possibile introdurre globalmente come legame geometrico la metrica canonica data dalla seguente espressione generale:

$$(1) \qquad ds^2 = d\tau^2 + 2g_{\alpha 4}(\xi^\alpha, \tau)\, d\xi^\alpha d\tau + g_{\alpha\beta}(\xi^\alpha, \tau)\, d\xi^\alpha d\xi^\beta \ ,$$

dove per semplicità si è posta la velocità della luce $c = 1$, mentre α, $\beta = 1$, 2, 3 sono gli indici di somma; $g_{ij}(\tau, \xi^\alpha)$ sono le componenti del tensore fondamentale, la coordinata τ, infine, rappresenta il tempo globale proprio della famiglia di linee di universo K con un differenziale $d\tau = d\xi^4$ definito in ogni punto delle linee di universo della famiglia K in uno spazio riemanniano avente per metrica la forma canonica (1).

La famiglia considerata K e le corrispondenti componenti del tensore fondamentale g_{ij}, in particolare, possono corrispondere a linee di universo del moto libero di singole particelle di pulviscolo non soggette a collisioni.

Partendo dalla definizione generale di accelerazione, in qualsiasi spazio

e in qualsiasi sistema di riferimento è facile dimostrare che l'accelerazione assoluta[1] $\mathbf{a}_{ab}$ delle particelle nelle coordinate τ, ξ^α per la metrica (1) è definita dalla formula

$$\mathbf{a}_{ab} = \frac{\partial g_{\alpha 4}(\tau, \xi^\alpha)}{\partial \tau} \ni^\alpha , \tag{2}$$

dove $\ni^\alpha$ sono i vettori della base del sistema di riferimento che corrisponde al sistema τ, ξ^α e alla metrica (1).

Come postulato fondamentale, suggerito dall'esperienza, ammetteremo che, come nella meccanica newtoniana, anche nelle meccaniche relativistiche, per ogni particella con massa costante, che porti nel suo moto libero un accelerometro, nel vuoto si avrà una condizione di imponderabilità, evidenziata dal fatto che la lancetta dell'accelerometro rimarrà immobile. Ciò significa che, qualunque sia il modello di spazio e tempo prescelto e fissato, sia nella meccanica newtoniana sia nelle teorie relativistiche, senza intervento di una qualsiasi forza esterna, fatta eccezione per quella gravitazionale, per ogni particella con massa costante m nel suo moto libero devono essere verificate le equazioni

$$-m\mathbf{a}_{ab} + m\mathbf{g} = \mathbf{0} \quad \text{ovvero} \quad \mathbf{a}_{ab} = \mathbf{g} , \tag{3}$$

dove $\mathbf{g}$ è il vettore accelerazione della forza gravitazionale considerato come funzionale dello spazio-tempo e della legge di movimento, mentre in ogni problema concreto per il movimento di particelle sperimentali è considerato come funzione delle coordinate dello spazio-tempo[2].

Le equazioni (3) sono corrette per ogni singola particella infinitesima che si muova liberamente e costituiscono le equazioni generali della meccanica dei corpi celesti.

In tal modo, nella teoria che esporremo più avanti, la simulazione del problema di un moto di singoli punti materiali con massa costante, o di movimento di polvere intesa come un continuo, è legata all'introduzione, per lo spazio-tempo quadridimensionale, della metrica (1) per la famiglia di linee di universo K e, in generale, di un vettore accelera-

(1) $\mathbf{a}_{ab}$ è calcolato in ogni punto delle linee di universo K nei sistemi inerziali in qualsiasi spazio quadridimensionale pseudoriemanniano.

(2) Ricordiamo che la presenza dell'accelerazione $\mathbf{g}$ non infrange le condizioni di imponderabilità della (3) e che è possibile considerare la forza di inerzia come la forza esterna della reazione spazio-tempo introdotta come un legame esterno definito dalla metrica.

zione $\mathbf{g}$ variabile, legato alla presenza di forze gravitazionali conformemente al postulato fondamentale (3).

Nella meccanica newtoniana il campo di accelerazioni $\mathbf{g}$ è introdotto e spiegato dalla legge di gravitazione universale di Newton. Tale spiegazione, tuttavia, parlando in generale da un punto di vista fisico, è inaccettabile, poiché contraddice l'invariabilità dei fenomeni gravitazionali rispetto alle trasformazioni di Lorentz e in connessione alla diffusione istantanea e protratta delle perturbazioni.

Tuttavia, se consideriamo il campo di accelerazioni $\mathbf{g}$ secondo la legge di gravitazione universale di Newton, in pratica è possibile calcolarlo. D'altro lato, gli esperimenti dimostrano che le leggi del movimento dei corpi celesti, secondo calcoli newtoniani, concordano assai bene con le osservazioni condotte nella realtà. Questo avviene perché nella pratica, in tutte le possibili simulazioni di routine, il rapporto v^2/c^2 è estremamente piccolo. In rapporto a ciò, una volta sostituito v^2/c^2 con zero nel passaggio dal sistema di riferimento delle particelle al sistema di riferimento dell'osservatore grazie alle trasformazioni di Lorentz, vengono esclusi i fondamentali effetti relativistici. I possibili errori locali, generati da questa scorrettezza fisica, saranno trascurabili. Se sulle linee di universo delle particelle in spazi riemanniani compaiono degli intervalli in cui non è possibile assumere v^2/c^2 uguale a zero, è possibile definire un campo preciso di accelerazioni gravitazionali, tenendo conto della teoria della relatività generale.

Ciò comporta la necessità di utilizzare, in luogo della legge di Newton sulla gravitazione universale, delle ipotesi supplementari che consentano una verifica a livello sperimentale.

Un esempio di postulato di questo genere è ammettere, come avviene nella teoria della relatività generale, che sia $\mathbf{g}=\mathbf{0}$, assunzione detta altrimenti condizione di assenza di forze gravitazionali su una curva di uno spazio adatto; pertanto, nella teoria della relatività generale, secondo il postulato (3) e la formula (2), si ha che $\mathbf{a}_{ab}=\mathbf{0}$ nei punti di qualsiasi linea d'universo delle particelle test. Perciò nella teoria della relatività generale tutte le linee d'universo di particelle che si muovano liberamente devono essere geodetiche.

È evidente che nella teoria della relatività ristretta e in spazi definiti occorre assolutamente introdurre un campo di accelerazioni gravitazionali $\mathbf{g} \neq \mathbf{0}$ e una forza di gravità $\mathbf{G}=m\mathbf{g} \neq \mathbf{0}$. Con tutto ciò, per spazi curvi pseudoriemanniani, con una definizione particolare nella teoria della relatività generale, la condizione $\mathbf{g}=\mathbf{0}$ può essere accettata in alcuni modelli di grande importanza per opportuni moti di particelle in regioni

dello spazio-tempo di dimensioni considerevoli(3). Con ciò si rende possibile sostituire gli effetti dell'attrazione gravitazionale con la curvatura di uno spazio adatto e sostituire il tempo assoluto con il valore del tempo globale nella (1) per la corrispondente famiglia K.

Nel caso generale, con opportune distinzioni, è possibile misurare in fase di esperimento il campo del vettore **g** o il campo dei suoi gradienti. Come risultato si otterrà una base sperimentale per la sua definizione che in generale è indipendente dalla legge di gravitazione universale di Newton o dalla teoria della relatività generale per oggetti fisici che posseggono una massa.

Varrà la pena ricordare che stiamo parlando della sostituzione della legge di gravitazione universale delle masse nella meccanica newtoniana con le leggi relativistiche di interazione tra singole particelle, la cui differenza da quelle newtoniane sorge per le grandi velocità tridimensionali delle particelle e per la curvatura dello spazio quadridimensionale.

Quando in sostituzione del modello newtoniano di spazio-tempo si sceglie un qualsiasi spazio pseudoriemanniano e quadridimensionale, per esempio lo spazio di Minkowski nella teoria della relatività ristretta o, in generale, un qualsiasi altro spazio concreto di Riemann, è possibile definire le corrispondenti proprietà del campo di accelerazioni **g** e della forza di gravità **G** in analogia con la fisica di Newton, ad esempio ricalcolando il campo, verificato sperimentalmente dalle accelerazioni tridi-

(3) Nel caso generale, tuttavia, non possiamo ignorare l'esperienza immediata che testimonia la presenza di forze gravitazionali. Il migliore esempio di ciò è l'osservazione di Newton della mela caduta dall'albero. È altresì evidente che, conformemente alla legge di gravitazione universale delle masse, nella meccanica newtoniana rientra la forza peso **G**, generata dall'accelerazione **g** ed è un fatto sperimentale comprovato.

D'altro lato, com'è noto, le equazioni della teoria del campo nella teoria della relatività generale, per il caso della materia disgregata (pulviscolo) con densità ϱ, in qualsiasi sistema di riferimento hanno la forma

$$R^{ij} - \frac{1}{2} g^{ij} R = \varkappa \varrho u^i u^j \ , \tag{*}$$

dove $\varkappa$ è un coefficiente numerico costante. In base alle identità di Bianchi, dalla (*) si ricava:

$$u^k \nabla_i (\varrho u^i) + \varrho u^i \nabla_i u^k = 0 \ .$$

Da qui deriva che $\nabla_i (\varrho u^i) = 0$ per la massa costante delle particelle, mentre il secondo termine, secondo la (3), dà:

$$u^i \nabla_i u^k = \frac{du^k}{d\tau} = g^k = 0.$$

In questo modo dalla (*) consegue che $\mathbf{g} = \mathbf{0}$ in qualsiasi punto regolare dello spazio per qualsiasi valore di ϱ.

mensionali nella meccanica newtoniana che vengono definite grazie alla legge di gravitazione universale o con la diretta misurazione delle accelerazioni o dei loro gradienti in esperimenti con particelle in libero movimento.

Indicheremo il metodo che, dato [4] un campo di accelerazioni $\mathbf{g}^* \neq \mathbf{0}$ in un qualsiasi spazio pseudoriemanniano R^*, permette di definire il campo di accelerazioni $\mathbf{g}$ per un problema analogo relativo al movimento di un punto materiale dal punto di vista di un osservatore arbitrario in R^* o in un altro spazio $R \neq R^*$ pseudoriemanniano.

Come problema analogo intendiamo dire che i sistemi di coordinate ξ^1, ξ^2, ξ^3 e le famiglie di linee K sono identici, mentre i tempi globali e le componenti metriche g_{ij} possono essere differenti restando però immutate le linee di universo $\xi^\alpha = \xi^\alpha_0 = \text{cost.}$ che riempiono l'intero spazio.

Faremo notare che, in questo caso, in ognuno degli spazi R^* e R i tensori fondamentali g^*_{ij} e g_{ij} nelle coordinate lagrangiane possono essere riportati alla forma canonica (1), da cui deriva, in generale

$$d\tau^* \neq d\tau, \quad \text{se} \quad R^* \neq R \,. \tag{4}$$

Se $g^*_{ij} = g_{ij}$, allora $R^* = R$; tuttavia in questo caso, accanto al sistema di coordinate ξ^α, τ, è possibile introdurre anche il sistema di coordinate dell'osservatore x^α, t con $x^\alpha \neq \xi^\alpha$ e $dt \neq d\tau$, ma possiamo comunque trovare le funzioni

$$x^k = x^k(\xi^1, \xi^2, \xi^3, \xi^4 = \tau), \qquad (k = 1, 2, 3, 4). \tag{5}$$

Inoltre, secondo la (5), dall'invariabilità della metrica ds^2 con $R^* = R$ si ricaverà che $g^*_{ij}(\xi^k) \neq g_{ij}(x^k)$ dove

$$g^*_{ij}(\xi) = g_{pq}(x) \frac{\partial x^p}{\partial \xi^i} \frac{\partial x^q}{\partial \xi^\sigma} \,. \tag{6}$$

Con $R^* = R$ la definizione delle quattro funzioni (5), altrimenti detta legge del moto continuo dei punti nel passaggio dalle variabili ξ^k alle variabili x^k, costituisce un problema di navigazione inerziale. Com'è

---

(4) Poiché i vettori dati $\mathbf{g}$ devono sempre essere confermati dall'esperimento. per stabilire un campo $\mathbf{g}$ con $v^2/c^2 \sim 1$ occorrono ancora delle ipotesi supplementari.

noto, esso si può risolvere sperimentalmente oppure con passaggi teorici per il calcolo di ξ^k e x^k, allorché sia dato $R^*=R$. In questo caso le trasformazioni (5) rappresentano una generalizzazione della trasformazione di Lorentz. È evidente che il passaggio da una trasformazione all'altra, dato uno spazio definito R^*, porta alla trasformazione in ogni punto dello spazio del vettore **g** nel vettore **g*** per la metrica stabilita conformemente alla formula tensoriale (6).

Se in uno stesso spazio di Riemann è ricavata una certa determinata soluzione del problema del movimento della materia disgregata nelle variabili x^1, x^2, x^3, $x^4=t$ dipendentemente dalle variabili lagrangiane η^1, η^2, η^3, $\tau=\eta^4$, ciò definirà la trasformazione quadridimensionale (5) nella forma:

$$x^\alpha=\varphi^\alpha(\eta^\alpha, t), \quad x^4=t \;. \tag{7}$$

Risulta qui evidente che le formule (7) determinano la legge del movimento $x^i(\eta^k)$ e la metrica in uno stesso spazio R, mentre le componenti tridimensionali della velocità v^α ricavate dalla trasformazione (7) possono essere anche considerate come funzioni delle coordinate η^α e t le quali, tuttavia, parlando in generale, non saranno canoniche secondo la definizione (1).

Se le variabili x^i vengono scelte nello stesso riferimento cartesiano, secondo Newton e secondo Minkowski nella teoria della relatività ristretta, risulta allora evidente che è possibile introdurre negli spazi di Newton e di Minkowski degli osservatori per i quali la descrizione dei movimenti sarà determinata dalle medesime funzioni

$$x^\alpha=\varphi^\alpha(\eta^\alpha, t) \;, \tag{8}$$

dove t è il tempo assoluto secondo Newton, che, tuttavia, non è uguale al tempo proprio τ secondo Minkowski, mentre le coordinate lagrangiane $\eta^\alpha=\text{cost.}$, t definiscono la medesima legge concreta di moto in entrambi i casi per differenti osservatori. Nella risoluzione del problema sul moto delle particelle è possibile adottare la metrica secondo Newton

$$dl^2=dx^{1^2}+dx^{2^2}+dx^{3^2} \;, \tag{9}$$

e secondo Minkowski

$$(9') \qquad ds^2 = dt^2 - dx^{1^2} - dx^{2^2} - dx^{3^2} = dt^2 - dl^2$$

e la legge di moto (8)[5].

In base alla trasformazione delle coordinate (7), che non risulta una trasformazione di Lorentz con il passaggio da x^i a η^k, possiamo scrivere

$$(10) \qquad ds^2 = (1 - v^2)(dt)^2 + 2(\ni_\alpha \cdot \mathbf{v})\, d\eta^\alpha dt + (\ni_\alpha \cdot \ni_\beta)\, d\eta^\alpha d\eta^\beta ,$$

dove $\ni_\alpha$ sono i vettori covarianti della base nel sistema η^k. Qui dt è il differenziale del tempo assoluto secondo Newton che coincide con il tempo dell'osservatore nella teoria della relatività ristretta. La formula (10) offre la metrica nel sistema di coordinate η^α, τ nello spazio di Minkowski. Per ricavare la metrica nel sistema canonico è sufficiente effettuare un'ulteriore trasformazione quadridimensionale

$$(11) \qquad \xi^\alpha = \eta^\alpha, \quad d\tau = dt\,\sqrt{1 - v^2} ,$$

dopo di che ricaveremo la metrica nella forma canonica

$$(12) \qquad ds^2 = d\tau^2 + 2g_{\alpha 4}\, d\xi^\alpha d\tau + g_{\alpha\beta}\, d\xi^\alpha d\xi^\beta ,$$

dove $g_{\alpha 4} = v_\alpha \sqrt{1 - v^2} = u_\alpha$, e pertanto è soddisfatta la formula (2), mentre $d\tau$ è l'elemento di tempo globale.

Se nel caso generale si ha lo spazio $R^* \neq R$, resta possibile introdurre una stessa famiglia di linee di universo K, la trasformazione (5) e le diverse componenti tensoriali delle metriche g^*_{ij} in R^* e g_{ij} in R; tuttavia si perderà il legame (6) tra questi tensori metrici.

Conformemente a ciò, è possibile definire lungo medesime linee K singole caratteristiche locali, cinematiche e tensoriali, ad esempio per le velocità quadridimensionali e tridimensionali $\mathbf{u}$ e $\mathbf{v}$ e per l'accelerazione $\mathbf{a}_{ab}$, *in* R^* grazie a g^*_{ij} e in R grazie a g_{ij}, basandoci sulla formula canonica (1).

È altresì evidente che i diversi vettori e le correlazioni tra essi, definite in R^*, e quindi gli analoghi vettori ugualmente definiti nello stesso senso in R, possono essere considerati anche in R^*, ma con un senso cambiato conformemente.

Tale condizione permette di passare dalle correlazioni e dalle caratteristiche in un spazio di Riemann alle correlazioni e alle caratteristiche

(5) I successivi passaggi sono basati sulla (8), che, in generale, è ricavata da diverse impostazioni del problema sul moto del pulviscolo.

in un altro spazio di Riemann. In particolare è possibile introdurre $\mathbf{a}^*_{ab}$ *in* R^* e $\mathbf{a}_{ab}$ *in* R lungo medesime linee $\xi^\alpha = \text{cost.}$ Resta quindi evidente che, secondo le condizioni di imponderabilità (3), ricaveremo che accanto alle equazioni $\mathbf{a}^*_{ab} = \mathbf{g}^*$ in R^* devono essere soddisfatte anche le equazioni $\mathbf{a}_{ab} = \mathbf{g}$ in R e risulterà inoltre(6) che $\mathbf{g}^* \neq \mathbf{g}$ e $\mathbf{a}^*_{ab} \neq \mathbf{a}_{ab}$.

In questo modo, dal campo di accelerazioni in R^* troviamo il campo di accelerazioni in R. Ad esempio, da un campo di accelerazioni nella teoria della relatività ristretta possiamo trovare un campo di accelerazioni in un qualsiasi spazio di Riemann, e pertanto negli spazi di Riemann per $K = K'$ la metrica e le componenti delle accelerazioni a_α e a'_α saranno differenti.

Detto ciò, occorre sottolineare che le componenti del campo di accelerazioni a_α in uno spazio $R^* = R$ si trasformeranno per le diverse $K(\eta^\alpha \to \xi^\alpha)$ come componenti dei medesimi vettori tridimensionali $\mathbf{a} = \mathbf{g}$, mentre nel passaggio da uno spazio all'altro essi cambiano.

Per una stessa famiglia di linee di universo K è possibile considerare diversi spazi riemanniani R^* e R e accelerazioni $\mathbf{g}^*$ e $\mathbf{g}$ se sono state scelte o assegnate le trasformazioni (5).

D'altro lato, se lo spazio R^* è stato fissato, allora in ogni suo punto, grazie alla navigazione inerziale, è possibile trovare per diversi osservatori le trasformazioni (5) e le corrispondenti famiglie K^* e K' nonché le corrispondenti componenti del vettore accelerazione per uno stesso vettore $\mathbf{g}$.

In particolare, se la metrica per la famiglia $\xi^\alpha = \xi^\alpha_0 = \text{cost.}$ nel sistema di coordinate lagrangiane ha la forma

$$ds^2 = d\tau^2 + 2g_{\alpha 4}(\xi^\alpha)\, d\tau\, d\xi^\alpha + g_{\alpha\beta}(\xi^\alpha, \tau)\, d\xi^\alpha\, d\xi^\beta, \tag{13}$$

dove τ è il tempo proprio, allora in ogni punto di questi spazi troveremo che

$$\mathbf{a} = \mathbf{g} = \mathbf{0}\ .$$

Dato che qualunque siano le trasformazioni delle coordinate in uno

---

(6) Come esempi particolari delle diverse metriche canoniche della forma (1) con $K = K'$, ma $R^* \neq R$, possiamo dare:

$$ds^2 = d\tau^2 + 2\,kg_{\alpha 4}\, d\xi^\alpha\, d\tau + g_{\alpha\beta}\, d\xi^\alpha\, d\xi^\beta\ ,$$

dove $k > 0$ è uno scalare costante arbitrario. Si osserverà facilmente che in questo caso per le componenti di accelerazione lungo linee di universo K varranno le formule $a'_\alpha = k\, a_\alpha$.

spazio stabilito, le geodetiche si trasformano sempre in geodetiche, risulta evidente che le forme generali delle metriche (1) e (13) sono invariabili.

Faremo ancora notare che per un qualsiasi spazio dato pseudoriemanniano in un sistema lagrangiano di coordinate (1) è sempre facile calcolare il vettore $\mathbf{a}_{ab}$ e in questo modo, sulla base della (3), ricaveremo il campo per il vettore $\mathbf{g}$, nel caso che questo campo sia noto in un qualche spazio dato, ad esempio nella teoria della relatività ristretta.

È evidente che soltanto dalle condizioni di imponderabilità (3) e dall'assenza di fatto di forze esterne di natura gravitazionale, in un moto libero, è impossibile stabilire in maniera univoca la metrica (1), il che è collegato alla necessità di scegliere un modello di spazio-tempo relativistico come pure un campo di accelerazione $\mathbf{g}$ per le forze di gravità. In rapporto a ciò possiamo porre il problema relativo alla scelta di condizioni supplementari per stabilire metriche di simulazione tra spazi quadridimensionali pseudoriemanniani.

In particolare, possiamo facilmente indicare una vasta classe di spazi pseudoriemanniani in cui i movimenti di pulviscolo possono avvenire solamente nelle condizioni di $\mathbf{a}_{ab} = \mathbf{g} = \mathbf{0}$ che figurano nella teoria della relatività generale.

Per scegliere concreti modelli di spazio, oltre a offrire un campo gravitazionale di accelerazioni $\mathbf{g} \neq \mathbf{0}$ oppure $\mathbf{a}_{ab} \neq \mathbf{0}$, confermato dall'esperienza, occorre ancora basarsi su una serie di postulati supplementari connessi a funzionali fisicamente giustificati e ai metodi per ricavare le corrispondenti equazioni (queste possono essere equazioni del tipo di Hilbert-Einstein; a seconda dei diversi problemi concreti e della loro risoluzione i corrispondenti spazi di Riemann saranno diversi in una vastissima gamma).

Questa condizione complica fortemente la teoria. È evidente che la teoria di simulazione di gravità secondo Newton in uno spazio definito è semplice e descrive il fenomeno della gravitazione con una grande esattezza. È per questo che va considerato uno sforzo più che naturale l'idea di A. A. Logunov di costruire un'analoga teoria meccanica nella teoria della relatività ristretta. I risultati teorici esposti sopra evidenziano l'essenza fisica della gravitazione e il cammino per costruire nuovi modelli tenendo conto degli effetti relativistici, ricordando che in fisica tutti i modelli matematici ammettono un approfondimento e un perfezionamento e che la teoria della relatività generale offre a questo proposito precisi motivi.

L'univocità della risoluzione di problemi connessi al moto di particelle di polvere, in presenza di corrispondenti equazioni nella teoria della

relatività generale, è possibile ottenerla, nella pratica attuale, soltanto in esempi particolari e solamente grazie a indispensabili ipotesi supplementari e particolari.